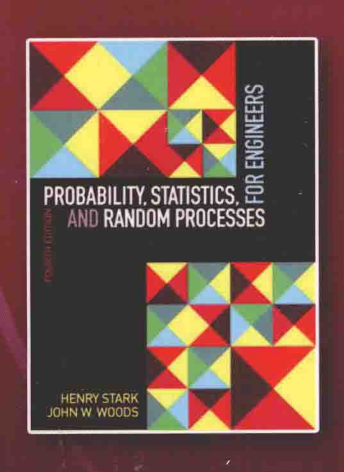
国外电子与通信教材系列

PEARSON

概率、统计与随机过程
（第四版）

Probability, Statistics, and Random Processes for Engineers
Fourth Edition



［美］ Henry Stark
John W. Woods 著

罗鹏飞 等译





概率、统计与随机过程（第四版）

Probability, Statistics, and Random Processes for Engineers

Fourth Edition

本书从工程应用的角度，全面阐述概率、统计与随机过程的基本理论及其应用。全书共9章，首先简单介绍概率论，然后各章分别讨论随机变量、随机变量的函数、均值与矩、随机向量、统计（包括参数估计和假设检验）、随机序列、随机过程基础知识，最后讨论了统计信号处理中的相关应用。书中给出了大量电子和信息系统相关实例，每章给出了丰富的习题。

本书适合作为电子信息类专业本科生和研究生的“随机信号分析”或“随机过程及其应用”课程的教材，也可供从事相关技术领域研究的科技人员参考。

相关图书



本书英文版

ISBN 978-7-121-17668-5

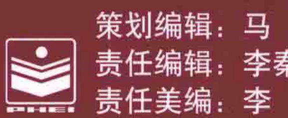
作者简介

Henry Stark 美国伊利诺伊理工学院荣誉教授，长期致力于概率、统计与随机过程的课程教学及其相关的科学研究工作。

John W. Woods 自1976年任教于纽约州特洛伊市的伦斯勒理工学院，长期讲授信息理论、随机过程、数字图像和视频处理等课程。研究兴趣包括图像和视频估计、复原、过滤和视频压缩编码。Woods博士是IEEE会士，兼任多家专业委员会的主席，曾因出色的科研和教学工作多次获奖。



PEARSON
www.pearson.com



策划编辑：马　岚
责任编辑：李秦华
责任美编：李　雯


ISBN 978-7-121-24805-4
9 787121 248054 >
定价：79.00 元

国外电子与通信教材系列

概率、统计与随机过程

（第四版）

Probability, Statistics, and Random Processes for Engineers

Fourth Edition

[美] Henry Stark
John W. Woods 著

罗鹏飞 等译

電子工業出版社
Publishing House of Electronics Industry

内容简介

本书从工程应用的角度，全面阐述概率、统计与随机过程的基本理论及其应用。全书共9章，首先简单介绍概率论，然后分别讨论随机变量、随机变量的函数、均值与矩、随机向量、统计(包括参数估计和假设检验)、随机序列、随机过程基础知识，最后讨论了统计信号处理中的相关应用。书中给出了大量电子和信息系统相关实例，每章给出了丰富的习题。另外有两章的补充内容，读者可从网站下载阅读。

本书适合作为电子信息类专业本科生和研究生的“随机信号分析”或“随机过程及其应用”课程的教材，也可供从事相关技术领域研究的科技人员参考。

图书在版编目(CIP)数据

概率、统计与随机过程：第4版/(美)斯塔克(Stark,H.)，(美)伍兹(Woods,J. W.)著；罗鹏飞等译.
北京：电子工业出版社，2015.4
(国外电子与通信教材系列)
书名原文：Probability, Statistics, and Random Processes for Engineers
ISBN 978-7-121-24805-4

Ⅰ. ①概…　Ⅱ. ①斯… ②伍… ③罗…　Ⅲ. ①概率论-高等学校-教材 ②数理统计-高等学校-教材 ③随机过程-高等学校-教材　Ⅳ. ①O21

中国版本图书馆 CIP 数据核字(2014)第271065号

策划编辑：马　岚
责任编辑：李秦华
印　　刷：涿州市京南印刷厂
装　　订：涿州市京南印刷厂
出版发行：电子工业出版社
　　　　　北京市海淀区万寿路173信箱　邮编　100036
开　　本：787×1092　1/16　　印张：31.5　　字数：976千字
版　　次：2015年4月第1次版(原著第4版)
印　　次：2015年4月第1次印刷
定　　价：79.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010)88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：(010)88258888。

序

2001年7月间，电子工业出版社的领导同志邀请各高校十几位通信领域方面的老师，商量引进国外教材问题。与会同志对出版社提出的计划十分赞同，大家认为，这对我国通信事业、特别是对高等院校通信学科的教学工作会很有好处。

教材建设是高校教学建设的主要内容之一。编写、出版一本好的教材，意味着开设了一门好的课程，甚至可能预示着一个崭新学科的诞生。20世纪40年代MIT林肯实验室出版的一套28本雷达丛书，对近代电子学科、特别是对雷达技术的推动作用，就是一个很好的例子。

我国领导部门对教材建设一直非常重视。20世纪80年代，在原教委教材编审委员会的领导下，汇集了高等院校几百位富有教学经验的专家，编写、出版了一大批教材；很多院校还根据学校的特点和需要，陆续编写了大量的讲义和参考书。这些教材对高校的教学工作发挥了极好的作用。近年来，随着教学改革不断深入和科学技术的飞速进步，有的教材内容已比较陈旧、落后，难以适应教学的要求，特别是在电子学和通信技术发展神速、可以讲是日新月异的今天，如何适应这种情况，更是一个必须认真考虑的问题。解决这个问题，除了依靠高校的老师和专家撰写新的符合要求的教科书外，引进和出版一些国外优秀电子与通信教材，尤其是有选择地引进一批英文原版教材，是会有好处的。

一年多来，电子工业出版社为此做了很多工作。他们成立了一个"国外电子与通信教材系列"项目组，选派了富有经验的业务骨干负责有关工作，收集了230余种通信教材和参考书的详细资料，调来了100余种原版教材样书，依靠由20余位专家组成的出版委员会，从中精选了40多种，内容丰富，覆盖了电路理论与应用、信号与系统、数字信号处理、微电子、通信系统、电磁场与微波等方面，既可作为通信专业本科生和研究生的教学用书，也可作为有关专业人员的参考材料。此外，这批教材，有的翻译为中文，还有部分教材直接影印出版，以供教师用英语直接授课。希望这些教材的引进和出版对高校通信教学和教材改革能起一定作用。

在这里，我还要感谢参加工作的各位教授、专家、老师与参加翻译、编辑和出版的同志们。各位专家认真负责、严谨细致、不辞辛劳、不怕琐碎和精益求精的态度，充分体现了中国教育工作者和出版工作者的良好美德。

随着我国经济建设的发展和科学技术的不断进步，对高校教学工作会不断提出新的要求和希望。我想，无论如何，要做好引进国外教材的工作，一定要联系我国的实际。教材和学术专著不同，既要注意科学性、学术性，也要重视可读性，要深入浅出，便于读者自学；引进的教材要适应高校教学改革的需要，针对目前一些教材内容较为陈旧的问题，有目的地引进一些先进的和正在发展中的交叉学科的参考书；要与国内出版的教材相配套，安排好出版英文原版教材和翻译教材的比例。我们努力使这套教材能尽量满足上述要求，希望它们能放在学生们的课桌上，发挥一定的作用。

最后，预祝"国外电子与通信教材系列"项目取得成功，为我国电子与通信教学和通信产业的发展培土施肥。也恳切希望读者能对这些书籍的不足之处、特别是翻译中存在的问题，提出意见和建议，以便再版时更正。

吴佑寿

中国工程院院士、清华大学教授

"国外电子与通信教材系列"出版委员会主任

出 版 说 明

进入 21 世纪以来，我国信息产业在生产和科研方面都大大加快了发展速度，并已成为国民经济发展的支柱产业之一。但是，与世界上其他信息产业发达的国家相比，我国在技术开发、教育培训等方面都还存在着较大的差距。特别是在加入 WTO 后的今天，我国信息产业面临着国外竞争对手的严峻挑战。

作为我国信息产业的专业科技出版社，我们始终关注着全球电子信息技术的发展方向，始终把引进国外优秀电子与通信信息技术教材和专业书籍放在我们工作的重要位置上。在 2000 年至 200 1 年间，我社先后从世界著名出版公司引进出版了 40 余种教材，形成了一套“国外计算机科学教材系列”，在全国高校以及科研部门中受到了欢迎和好评，得到了计算机领域的广大教师与科研工作者的充分肯定。

引进和出版一些国外优秀电子与通信教材，尤其是有选择地引进一批英文原版教材，将有助于我国信息产业培养具有国际竞争能力的技术人才，也将有助于我国国内在电子与通信教学工作中掌握和跟踪国际发展水平。根据国内信息产业的现状、教育部《关于“十五”期间普通高等教育教材建设与改革的意见》的指示精神以及高等院校老师们反映的各种意见，我们决定引进“国外电子与通信教材系列”，并随后开展了大量准备工作。此次引进的国外电子与通信教材均来自国际著名出版商，其中影印教材约占一半。教材内容涉及的学科方向包括电路理论与应用、信号与系统、数字信号处理、微电子、通信系统、电磁场与微波等，其中既有本科专业课程教材，也有研究生课程教材，以适应不同院系、不同专业、不同层次的师生对教材的需求，广大师生可自由选择和自由组合使用。我们还将与国外出版商一起，陆续推出一些教材的教学支持资料，为授课教师提供帮助。

此外，“国外电子与通信教材系列”的引进和出版工作得到了教育部高等教育司的大力支持和帮助，其中的部分引进教材已通过“教育部高等学校电子信息科学与工程类专业教学指导委员会”的审核，并得到教育部高等教育司的批准，纳入了“教育部高等教育司推荐——国外优秀信息科学与技术系列教学用书”。

为做好该系列教材的翻译工作，我们聘请了清华大学、北京大学、北京邮电大学、南京邮电大学、东南大学、西安交通大学、天津大学、西安电子科技大学、电子科技大学、中山大学、哈尔滨工业大学、西南交通大学等著名高校的教授和骨干教师参与教材的翻译和审校工作。许多教授在国内电子与通信专业领域享有较高的声望，具有丰富的教学经验，他们的渊博学识从根本上保证了教材的翻译质量和专业学术方面的严格与准确。我们在此对他们的辛勤工作与贡献表示衷心的感谢。此外，对于编辑的选择，我们达到了专业对口；对于从英文原书中发现的错误，我们通过与作者联络、从网上下载勘误表等方式，逐一进行了修订；同时，我们对审校、排版、印制质量进行了严格把关。

今后，我们将进一步加强同各高校教师的密切关系，努力引进更多的国外优秀教材和教学参考书，为我国电子与通信教材达到世界先进水平而努力。由于我们对国内外电子与通信教育的发展仍存在一些认识上的不足，在选题、翻译、出版等方面的工作中还有许多需要改进的地方，恳请广大师生和读者提出批评及建议。

电子工业出版社

教材出版委员会

译 者 序

Henry Stark 是美国伊利诺伊理工学院荣誉教授，长期致力于概率、统计与随机过程的课程教学及其相关的科学研究工作，该书自 1986 年以来已先后出版了四个版本，作者试图以浅显易懂、富有启发和有趣的方式为工科学生介绍概率、随机过程，允许主讲教师灵活地选取教学内容，除了概率、随机变量和随机过程课程中标准内容的介绍外，书中还涉及了建模、基本统计技术、计算机模拟、可靠性、熵、马尔可夫链和排队论等。

在电子与计算机工程中遇到的系统复杂性要求学生不仅对概率论与随机过程的概念有深刻的理解，而且也要求学生要有使用概率工具的能力，因此，入门性的课程不仅应该包含基本理论，而且也应该能解决工程中的实际问题，学生不仅要有解决问题的技能，也要理解如何将实际问题转化为问题求解的概率模型，理论与实际紧密结合。从实际问题到概率模型的转换方法有很多，首先，重要的概念通常都由展示的实际数据或计算机模拟数据建立起来；其次，书中选取了一部分基本的统计技术，这一部分展示了统计方法在理论与实际之间的桥梁作用。最后，重要的随机变量和随机过程都是通过建模时增加变量、从简单到复杂的方式建立起来。

启发和激励学生对于概率与随机过程的课程是一个很大的挑战，这要求教师为学生展示概率和随机过程的理论与工程实际的联系。书中选择了大量的与现实生活相联系的实例，如休闲娱乐中的游戏、选美、湖中鱼的数量估计，公共服务中的入住问题、等车时间、客户服务时间，可靠性理论中的设备故障率、备用冗余、多处理器的可靠性，图像处理中的图像压缩、边缘检测、医学图像的疾病判定，通信中的光电检测、通信误码率、分组交换、调制信号分析，信号处理中的模数转换器、线性预测、滤波器分析、内插与抽取，互联网中的网页排名等。书中还给出了大量的 MATLAB 程序，并给出了部分需要采用 MATLAB 才能完成的作业。

与教材配合的网站给出了本书的两章扩展内容①，第 10 章为概率和随机过程的高级话题，包括均方微积分、随机微分方程、遍历性、Karhunen-Loève 展开、带限过程和周期过程的表示等；第 11 章为统计信号处理中的应用，包括随机变量和随机向量的估计、新息序列和卡尔曼滤波、随机序列的维纳滤波、数学期望最大化(EM)算法、隐马尔可夫模型、谱估计、模拟退火算法等。网站上还包含了一部分演示概率论和随机过程概念的 MATLAB 程序②。

本书每一章都有大量的例题来展示问题的求解技术，通过问题简化来建立概念和说明其应用。书中包括了 700 多道习题，每一章的最后还增加了一些附加题，这些题需要阅读补充材料才能完成。在本书的网站上还给出了许多测试题和多项选择题，这是作者 1988 – 2006 年在伊利诺伊理工学院任教期间为高年级本科生和研究生出的试题。

参与本书翻译的有罗鹏飞(第 1 章、第 2 章)、张文明(第 6 章、第 7 章)、谢晓霞(第 8 章、第 9 章)、许可(第 3 章、第 4 章，第 5 章、附录 A ~ 附录 G)，最后由罗鹏飞对全书的译文进行了校对和整理。

由于译者的水平有限，文中难免有不当之处，敬请读者批评指正。另外，在本书即将付印之前，收到了原作者的勘误表。我们对其逐一进行了修订。

---

① 登录华信教育资源网(http://www.hxedu.com.cn)可下载本书第 10 章和第 11 章的内容(英文版)。

② 采用本书作为教材的教师可获得本书的配套教辅，请联系 010-88254555 或发邮件至 te_service@phei.com.cn 索取。

绪　　言

当前的版本相对于前一版本有了相当大的改变，作者试图保持通俗易懂的风格，少用测度理论，也就是比测度理论方法少一点数学，但又比一般公式和手册指南多一份严谨。

人们都说概率很难理解，不仅是因为数学基础，而是因为它产生了许多与直观不相符的结果。在许多应用型的学生中，概率论有许多批评论调，在这些批评论调中，居首位的就是“我们为什么需要概率论呢?”这种批评很容易回答，因为未来的工程师和科学家将会认识到，几乎人类的每一次尝试都包含在不确定或者概率的环境中做出决定，像保险、气象、城市规划、医药等领域确实都是如此。另一种更有说服力的批评是：“如果概率论提供的答案不是确定的，而只是一种推断和可能，那么它又好在哪里呢?”这里的答案就是，即使在不确定的领域里，也可以做许多好的计划和精确的预测。此外，应用概率——常常也称为统计学，确定提供了几乎确定的结果：在政治性民意测验和预测中就提供了有力的证据。

在前一版本中，我们只是稍微涉及了统计学的领域，更强调随机过程和信号处理领域。在本书的电子版中，研究生级的信号处理、随机过程的高级话题讨论与统计学方面的新材料一起被保留下来，而在本书中，我们删除了统计信号处理的应用、随机过程的高级话题以及模式识别入门资料的那些章节。

目前的版本用了很多说明性的例子和详细的讨论，尽了很大的努力使学生能够理解概念。我们把一些短语的应用降到最小程度，如“很容易证明……”、“可以证明……”、“容易看出……”，等等。我们尝试提供一些现实中的例子，如“药效”、“传染的可能性”、“赌博中赢的几率”，以及数字通信、网络和信号等。

另一个主要的变化是增加了统计学基础及其在现实中的应用两章内容，第一部分讨论参数估计，第二部分讨论假设检验。工程中有许多问题涉及参数估计，例如，从估计一个新方案的优点到估计计算机之间的交通信号的数量，此外，许多工程活动都需要在随机环境中做出判决，从决定一个新药是否有效，到判定一种新的教学方法是否有效。用详细的实例介绍和讨论了一些标准统计工具的来源和应用，如 t 检验、χ^2 检验和 F 检验等，在每章最后还给出了对应的习题。

最后，在本书的网站上还给出了许多测试题和多项选择题，这是其中一位作者于 1988 年至 2006 年间在伊利诺伊理工学院任教期间为高年级本科生和研究生出的试题。网站也包含了一部分演示概率论概念的 MATLAB 程序。

总而言之，熟悉第三版的读者将会看到下列明显的变化：

- 在统计学的分支方面增加了称为“参数估计”的新的一章，并给出了许多说明性的例题。
- 在统计学的分支方面增加了称为“假设检验”的新的一章，并给出了许多说明性的例题。
- 给出了大量新的习题，习题的难度不同，用以检验学生对统计学原理的掌握。
- 在网站上给出了大量的测试题和多项选择题，并在每一章给出了标准解答。

- 从概率与统计的应用领域选取了许多应用实例。
- 更多地借助计算机来辅助教学，例如，(i)概率现象的图形显示；(ii)利用 MATLAB 程序来说明概率论的一些概念；(iii)部分作业要求使用 MATLAB 或 Excel。
- 许多来自学生反馈的问题，在文中已经过修改，从而促进学生对教材难点问题的理解。

Henry Stark
John W. Woods

目　　录

网上章节(英文版)[①]

① 登录华信教育资源网(http:/www.hxedu.com.cn)可免费下载。

第1章　概率论导论

1.1　引言：为什么要学习概率论

初学概率论的学生问得最多的问题是：随机的本质是什么？真正的随机与由于信息缺乏而当成随机有何区别？首先考虑有关随机现象的问题。"随机现象存在吗?"当我们用天文望远镜观测宇宙时，可看到许多星系、恒星、行星组成的巨大阵列，以明显的随机顺序和位置排列。

从另一个极端来说，宇宙的尺寸在原子级又会怎样呢？我们的朋友，物理学家们把这样的事情说成是原子系统处于某种状态的概率，不确定性原理告诉我们，粒子的位置和动量能够同时被测定的精度是有限制的，两个量是模糊的和不确定的。

包括一些著名的物理学家在内的许多人都相信，自然界的本质是随机的。欧根(Eugen Merzbacher)在他的量子力学的著名教科书[1-1]中写道：

量子力学的概率理论认为，刚才那个例子所说的不确定性是自然界固有的特性，不仅仅是专业的暂时缺乏，需要在未来有更好和更完善的理论去解决。这一常规的解释否认了那种包含当前量子力学的、没有瑕疵的理想理论的可能性，量子力学最难听的词"不完美"就是对严格经典的确定性的抛弃。

然而，在讨论概率方法的合理性时，确定性与固有的不确定性问题是永远也不需要考虑的。事实上有数不清的情况，在考虑一个现象时无法做出明确的结论，因为观测不到所有对该现象做出贡献的所有因素。例如，预测电阻 R 由热引起的噪声电流 $i(t)$，毫无疑问，如果持续跟踪大约 10^{23} 个受激电子在彼此磁场中的移动并产生本地场脉冲，这些本地场脉冲最终对产生电流做出贡献，就能精确地预测未来某个时刻的电流 $i(t)$，这样的计算是不可思议的，因此使用概率模型而不是麦克斯韦方程来处理电阻噪声。类似的话题还有气象预测、投掷一枚真实硬币的结果、计算机失效时间、CMOS 成像器的暗电流等。我们得出结论：确定性与非确定性，无论你站在哪一边，在现实生活中都被迫采用概率模型，因为我们不知道，也不可能计算和测量所有对事件做出贡献的力，这种力太复杂、太多、也太微弱。

概率是一种帮助我们用平均研究物理系统的数学模型，我们应该能够在相同条件下重复实验。因此，诸如"彗星明天撞击地球的可能性是多少?"或者"其他行星上存在生命的可能性有多大?"，这样的问题绝不要用概率来回答，这里的问题是我们没有过去的类似的"实验"数据。

费希尔(R. A. Fisher)和米泽斯(R. von Mises)在20世纪早期对建立现代概率论的基本理论做出了重要贡献，本书许多现代公理化的处理大都是依据科尔莫戈罗夫[1-2](Andrei N. Kolmogorov)工作的结果。

1.2　概率的不同类型

有四种基本的概率类型，这里将简要进行讨论。

直观概率

这类概率是基于直观来处理和判断的。这样，“她可能与他结婚”和“他可能开得太快”就属于这一类。直观概率会得出一些矛盾的特性。乔(Joe)仍有可能买进口的 Itsibitsi，因为它的可靠性在全球是著名的，即使他的邻居弗兰克(Frank)有一辆使用了 19 年从未抛过锚的别克汽车，而且乔的另一个邻居比尔(Bill)在修理店也有一辆 Itsibitsi。这样，乔自然会通过统计进行“合理的”推断，而忽略了它的个人观测。另一方面，如果弗兰克说比尔认为他或许在当地的池塘看见了鳄鱼，那么，乔就会对他九岁女儿在那个池塘游泳担忧。尽管事实是没有人说曾经在这个池塘看见过鳄鱼，无数的人在那个池塘游泳也没有被鳄鱼咬过。为了给这个例子以某种可信度，假定池塘是在佛罗里达州。这里，乔忽略了统计而对谣传做出了反应。为什么？可能是因为池塘里有鳄鱼的“意外事情”对乔的代价太高[1-3]。

购买彩票的人直观地相信某些数的组合，如他们孙子的生日，要比 06-06-06 更容易中奖。很多人会根据在之前的行为下注，如果一个硬币在前 7 次的投掷都是正面，那么下一次投掷是否仍为正面？许多人都认为硬币有某种记忆，因此，在 7 次正面后，在接下来的投掷中硬币出现更多的反面是很正常的事情。

处理直观概率的数学理论是由库普曼(B. O. Koopman)建立的[1-4]，但在本书不讨论这个内容。

古典概率

在这一方法中，事件的概率不是实验性的，而是通过预先计算事件 E 可能发生的次数 n_E，形成一个比值 n_E/n，其中 n 是所有可能的结果，也就是事件 E 发生的次数与其他次数的总和 n_E。这里应注意的是所有结果是等可能的。假定掷一对无偏的六面骰子，问得到 7 的概率是多少。我们将结果空间划分为 36 个等可能的结果，如表 1.2-1 所示，表中每一个结果是两个骰子数之和。

表 1.2-1 投掷的结果

两个骰子

第二个骰子	第一个骰子					
	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

投掷结果总数是 36，得到 7 的次数是 6。因此

$$P[\text{得到 7}]=\frac{6}{36}=\frac{1}{6}$$

例 1.2-1 (掷一枚均匀硬币两次)可能的结果是 HH, HT, TH 和 TT，至少得到一次反面的概率计算如下：用 E 表示至少得到一次反面的事件，则 E 是下列结果的集合：

$$E=\{\mathrm{HT},\mathrm{TH},\mathrm{TT}\}$$

这样，无论结果是 HT 或者 TH 或 TT 事件，事件 E 都发生。事件 E 元素的个数是 $n_E=3$，实

验结果总数 N 是 4，因此

$$P[\text{至少一次反面}]=\frac{n_E}{n}=\frac{3}{4}$$

注意，由于没有涉及物理实验，假定理想的"均匀硬币"是没有问题的，在古典概率中，每一个实验都没认为是"公平"的。

古典概率至少会遇到两个问题：(1)它不能处理实验结果不是等可能的情况；(2)它不能处理实验结果是无穷大的情况，即 $n=\infty$ 的情况。然而，在那些不可能通过实验去确定实验结果概率的情况下，以及确实对等可能实验结果有争论的情况下，古典概率是有用的。

从历史上看，古典概率的前身是米泽斯[1-6]在 20 世纪 30 年代建立的相对频率的方法。

频率作为概率的测度

相对频率的方法定义事件 E 的概率要进行 n 次重复实验，事件 E 发生的次数用 n_E 表示，那么，事件 E 发生的概率定义为

$$P[E]=\lim_{n\to\infty}\frac{n_E}{n} \tag{1.2-1}$$

很显然，由于 $n_E\leqslant n$，所以，$0\leqslant P[E]\leqslant 1$。这一方法的困难在于我们不可能做无穷多次实验，所以，只能从有限次实验去估计 $P[E]$。其次，假定当 n 趋向无穷时 n_E/n 趋向于一个极限，但是，考虑投掷 1000 次均匀硬币，精确地得到 500 次正面的可能性是非常小的；事实上，如果投掷 10 000 次，得到 5000 次正面的可能性甚至会更小。当 $n\to\infty$ 时，精确地观测到 $n/2$ 次正面的事件变得非常小，但是，对于均匀硬币，我们的直观要求 $P[\text{正面}]=1/2$。假定选择一个 $\delta>0$，如果硬币确实是均匀的，那么，通过实验发现，随着 n 变得很大

$$\left|\frac{n_E}{n}-\frac{1}{2}\right|>\delta \tag{1.2-2}$$

的次数将变得很小。这样，尽管在实验的任何阶段，特别是当 n 很大时，n_E/n 精确地等于 1/2 是不可能的，这个比值永远都在 1/2 上下徘徊。根据式(1.2-2)，偏离 1/2 附近的次数确实变得很小。

用相对频率定义概率尽快存在这些，但在将概率论应用于现实世界时，相对频率的概念仍是很重要的。

例 1.2-2　(random. org)互联网的一个随机数源是 RANDOM.ORG，这是由爱尔兰都柏林的圣三一学院的计算机与统计学院的一个教授建立的，他依据大气噪声的函数来计算随机数，并且通过了随机性真实的许多统计检验。利用该网站的免费服务，下载了 10 000 个随机数，每一个随机数等可能地在 1～100 取值，写了一个 MATLAB 函数 RelativeFrequencies()，利用该函数读取随机数文件并画出比值 n_E/n 与实验次数 n 的曲线，$n=1,\cdots,10\ 000$。选择事件 E 表示 100 个数中任意一个数的出现。例如 $E\triangleq\{\text{数字 5 出现}\}$。次数 n_E 表示到 n 次实验中数字 5 出现的次数，得出的输出如图 1.2-1 所示，图中可以看出，对于 100 个等可能的数，它有收敛到理想值 0.01＝1/100 这样一个趋势。图 1.2-2 给出了数字选择为 23 时的图形，它同样也有收敛到 0.01 的趋势。在两种情况下，我们注意到对于任意的 n，尽管这种收敛值并不是精确的，但是它还是收敛到理想值的附近。

这个程序放在本书的网站上。

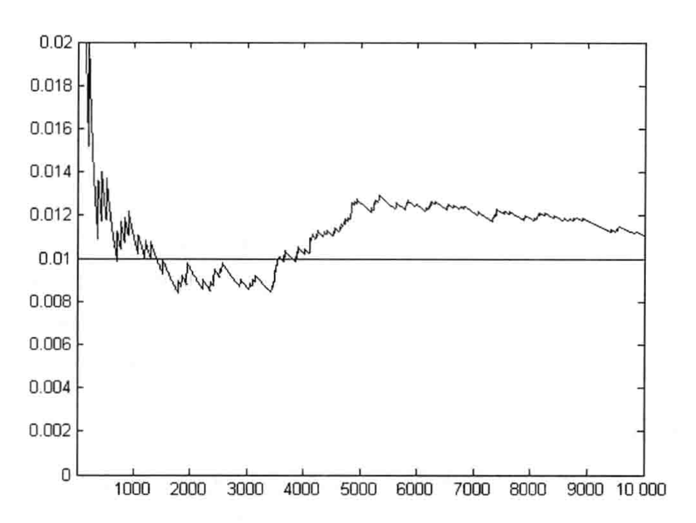


图 1.2-1 从大气噪声画出的 $E=\{$数字 5 出现$\}$的 $n_E/n\sim n$ 图(引自网站 RANDOM.ORG)

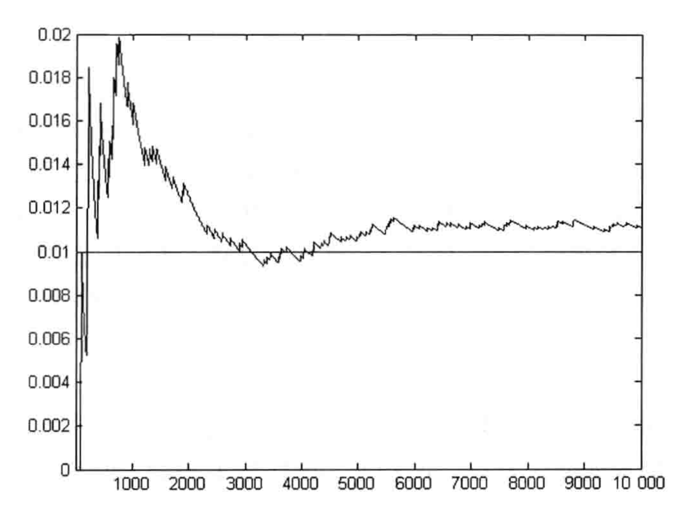


图 1.2-2 从大气噪声画出的 $E=\{$数字 23 出现$\}$的 $n_E/n\sim n$ 图(引自网站 RANDOM.ORG)

基于公理化理论的概率

公理化方法是出现在大多数近代的教材中。为了建立这一理论，我们必须引入某些概念，特别是诸如随机实验、样本空间、事件。简单地说，随机实验是一种实验结果带有不确定性的实验，在每次实验中，实验的可能结果不止一个，这也是随机实验中“随机”这个词的由来；样本空间是随机实验所得结果的集合；事件是样本空间中满足某些约束的子集，例如，我们希望能够计算每个事件的概率。此外，在不可数或者连续的样本空间的情况下，对那些有技术约束的子集也叫做事件；只有一个实验结果的事件称为基本事件。这些术语将在 1.4 节和1.5 节进行更为详细的解释。

1.3 概率的误用、误算和悖论

在日常生活中概率的误用是很常见的，下面给出一些误用的例子。考虑一个谋杀案的被告，对谋杀他妻子一案并不认罪，被告打他妻子有很大的偶然性。他的律师辩护说，被告确实

打过妻子，但是很多男人都打老婆，在打老婆的男人中谋杀老婆的概率只有 0.001，即千分之一。让我们假定这一陈述是真实的话，那就意味着通过隐喻打老婆并不表明谋杀老婆这样一个事实来影响陪审团。遗憾的是，除非陪审团成员学过概率论的课程，他们可能并没有意识到下面更有意义的问题：假定一个受虐待的老婆被谋杀，丈夫是凶手的概率是多少呢？事实上，统计表明这个概率大于 1/2。

在 1996 年的总统竞选中，参议员多尔(Bob Dole)的年龄成了大家关注的话题，他的对手声称 72 岁的男性白人在随后五年死去的风险有 27% 。这引起了争论，如果多尔当选，那么在他任期内去世的概率大于四分之一。争论的问题是，在计算生存概率时没有附加相关事实作为条件。即便如此，当一个 72 岁的男性仍在工作，且比较富有，那么考虑这些附加的事实，73 岁(假定多尔就任总统的岁数)平均来说在后四年死去只有八分之一的机会[1-3]。

概率的误用在对外星球生命的预测中也频繁出现，在奥采尔(Amir Aczel)的“Probability 1”(Harcourt Brace & Company，1998)一书中要我们确信外星人肯定存在，只等我们去发现。然而，伦敦帝国理工大学的约翰·多伦特(John Durant)为奥采尔书写了富有说服力的评述：

统计学十分强大和重要，奥采尔对此非常清楚。但是，关于宇宙的规律是不能用统计学去替代经验知识，我们还没有一种合理的方式来对宇宙的运行规律做出稳定的估计。我们也没有一种合理的方式得出在条件合适的情况下生命自然出现的概率的稳定估计。因此，除非我们发现了外星人之前或者深刻理解了至少一种陆地生命的起源，我们只能猜测在宇宙的其他地方存在生命的可能性，只要我们是在猜测，就不要对我们有趣的猜测披上数学确定性的外衣。

基于相对频率的概率计算有可能导致悖论，在参考文献[1 -3]中可以找到一个很好的例子，这里重述如下：

在一组 35 ~50 岁的美国妇女的样本中，有 4% 的人在一年内患乳腺癌。那么，49 岁的史密斯(Smith)女士在第二年患乳腺癌的机会达 4%？这个问题没有答案。假定一组 45 ~90 岁的妇女样本类，即史密斯女士所属的类，有 11% 的人在一年内患乳腺癌的，那么，史密斯患乳腺癌的机会是 4% 呢还是 11%？假定她的母亲患过乳腺癌，而母亲患过乳腺癌的 45 ~90 岁年龄段的妇女中，22% 的人会得乳腺癌，那么，史密斯患乳腺癌的机会究竟是 4%、11% 或 22%？她是希腊后裔，还抽烟，生活在加利福尼亚，25 岁前有两个孩子，40 岁以后又生了一个，我们该把她归到哪个类来算出患乳腺癌的机会呢？你或许会认为，她的类越具体，结果越准确。然后，类越具体，类的数量就越小，频率的可靠性也就越低。如果在世界上非常像史密斯的人只有两个，其中一个患乳腺癌，那是不是该说史密斯患乳腺癌的机会是 50%？在极限的情况下，唯一与史密斯的详细细节具有可比性的类就是史密斯自己，但在这一个类里，相对频率是没有意义的。

前面的例子应该不会留下这样的印象：基于相对频率的概率是无用的。第一，有许多工程和科学问题并不像史密斯女士患癌症的可能性那么复杂。第二，如果我们提炼类，从而减少类的容量，那么，基于相对频率的估计就会变得不稳定。但究竟有多大的不稳定是概率和统计的研究领域(参见 4.4 节的大数定理)。第三，有许多情况，为了得到好的概率估计，类的提炼所要求的条件是类的容量足够大。再回到史密斯女士的问题，如果类的容量变得太小，那么，应该停止添加条件，学会用大的类去估计概率，这类估计对于筛选实验也就够了。

1.4 集合、域和事件

集合是客观物体的总体，物体可以是具体的，也可以是抽象的。集合的一个例子就是纽约所有身高超过 6 英尺的居民的总体，集合的子集就是包含在一个更大集合中的一个总体。身高在 6 ~6.5 英尺的纽约居民的总体是前一个集合的子集。在概率论中，称集合为事件，我们特别感兴趣的是所有随机实验结果的集合及其子集。所有实验结果的集合用 Ω 表示，单一的实验结果用 ζ 表示。集合 Ω 称为随机实验的样本空间。Ω 的子集称为事件，事件发生的概率是我们感兴趣的。Ω 自身也称为必然事件，而空集 ϕ 也称为零事件。

样本空间的例子

例 1.4-1 (投掷硬币)随机试验由投掷一次硬币构成，那么 $\Omega=\{H, T\}$，其中 H 表示正面，T 表示反面。

例 1.4-2 (投掷硬币两次)随机试验由投掷硬币两次构成，那么 $\Omega=\{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$，其中 H 表示正面，T 表示反面。$\Omega$ 有 16 个子集，其中一个子集 $E=\{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}\}$ 是在两次投掷中至少有一次是正面的事件。

例 1.4-3 (头上的头发)实验是随机挑选一个人，数他(她)头上的头发，那么

$$\Omega=\{0,1,2,\cdots,10^7\}$$

即到 10^7 的所有非负整数的集合，这里假定了没有一个人的头发会超过 10^7 根。

例 1.4-4 (夫妻双方的年龄)随机实验由确定随机选定的一对年龄最接近的夫妻的年龄组成。用 x 表示男方的年龄，y 表示女方的年龄，Ω 可描述为

$$\Omega=\{\text{二元组}(x, y): x \text{ 是 } 10 \sim 200 \text{ 之间的任何整数}; y \text{ 是 } 10 \sim 200 \text{ 之间的任何整数}\}$$

注意到在例 1.4-4 中，假定没人能活过 200 岁，也没人在 10 岁前结婚。类似地，在例 1.4-1 中，硬币永远不会以边沿着地。如果后者是一种可能的结果，它必须包含在 Ω 中，因为 Ω 是所有实验结果的集合，它是必然事件。

例 1.4-5 (弹性碰撞的角度)随机实验是由观察弹性碰撞实验中核粒子的偏转角度构成的。那么

$$\Omega=\{\theta: -\pi \leqslant \theta \leqslant \pi\}$$

Ω 的一个子集的例子是

$$E=\left\{\frac{\pi}{4} \leqslant \theta \leqslant \frac{\pi}{4}\right\} \subset \Omega$$

例 1.4-6 (电力)随机实验是由测量一个电流驱动的电阻的瞬时功率组成的。那么

$$\Omega=\{P: P \geqslant 0\}$$

由于功率不可能为负的，我们忽略 P 的负值。Ω 的一个子集是事件 $E=\{P>10^{-3}\mathrm{W}\}$。

注意到在例 1.4-5 和例 1.4-6 中，在 Ω 中的元素数量是不可数无穷的，因此，它的子集也是不可数无穷的。当实验结果的数量是有限时，如例 1.4-4 所示，不同子集的实验结果的数量也是有限的，每一个子集表示一个事件。这样，如果 $\Omega=\{\zeta_1,\cdots,\zeta_N\}$，$\Omega$ 的可能子集数有 2^N

个，这一点可以通过考察 ζ_i 是不是在 Ω 任何一个子集中得出。这样可以得出 2^N 个不同的子集或事件，包括必然事件和不可能事件或零事件。

集合论回顾　两个集合 E 和 F 的并(和)记为 $E\cup F$ 或 $E+F$，它是由所有那些在 E 或 F 中的元素构成。因此，如果 $E=\{1,2,3,4\}$ 和 $F=\{1,3,4,5,6\}$①

$$E\cup F=\{1,2,3,4,5,6\}$$

如果 E 是 F 的子集，用 $E\subset F$ 表示，很显然，如果 $E\subset F$，可得出 $E\cup F=F$。用 ζ 表示 Ω 的元素，或“属于”Ω，记为 $\zeta\in\Omega$，这样，可以记为

$$E\cup F=\{\zeta\colon\zeta\in E\text{ or }\zeta\in F\}\tag{1.4-1}$$

其中“or”表示“或”。显然，$E\cup F=F\cup E$。两个集合 E 和 F 的交或积记为 $E\cap F$ 或 EF，它是那些同时在 E 和 F 中元素的集合。在前面的例子中

$$EF=\{1,3,4\}$$

正规的写法是 $EF\triangleq\{\zeta\colon\zeta\in E$ 和 $\zeta\in F\}=FE$，集合 E 的补集记为 E^c，它是由所有不在 E 中的元素的集合。由此可见，如果 Ω 是样本空间或更一般的全集，则

$$E\cup E^c=\Omega\tag{1.4-2}$$

另外，$EE^c=\phi$，集合的差或者是更一般地称为 E 减 F，它是由那些在 E 中但不在 F 中的元素组成，记为 $E-F$。很显然

$$E-F\triangleq EF^c$$

$$F-E\triangleq FE^c$$

但需要注意的是，集合的差不像数字的差，例如，$F-E-E=F-E$。两个集合的不相容(exclusive)是由在 E 或 F 中、但不同时在 E 和 F 中的那些元素构成，记为 $E\oplus F$。很容易证明②

$$E\oplus F=(E-F)\cup(F-E)\tag{1.4-3}$$

并、交等运算可以用维恩图说明，维恩图在辅助推理和建立概率关系是很有用的。集合的不同运算 $E\cup F$，EF，E^c，$E-F$，$F-E$，$E\oplus F$ 在图 1.4-1 中用斜线表示。

如果两个集合 $EF=\phi$，即它们没有共同元素，我们称 E 与 F 是互斥的。给定任何任意集合 E，E 的 n 划分是由一系列集合 E_i 构成，其中 $i=1,2,\cdots,n$，$E_i\subset E$，$\cup_{i=1}^{n}E_i=E$，且对于 $i\neq j$，有 $E_iE_j=\phi$。因此，给定两个集合 E 和 F，F 的 2 划分为

$$F=FE\cup FE^c\tag{1.4-4}$$

使用维恩图很容易得出下列结果：

$$(E\cup F)^c=E^cF^c\tag{1.4-5}$$

$$(EF)^c=E^c\cup F^c\tag{1.4-6}$$

由数学归纳法可以证明③：给定集合 $E_1,\cdots,E_n$，有

$$\left[\bigcup_{i=1}^{n}E_i\right]^c=\bigcap_{i=1}^{n}E_i^c\tag{1.4-7}$$

---

① 记住：在集合中元素的顺序并不重要。

② 式(1.4-3)表明并用 $\cup$ 而不用 $+$ 表示，刚学习的学生也可能错误地写 $(E-F)+(F-E)=E-F+F-E=0$，这个式子没有任何意义，注意 $F+F\neq 2F$，这个式子也是没有意义的，事实上，$F+F=F$。所以，在集合论中用 $+$ 或 $-$ 时要仔细一点。

③ 数学归纳法的含义请参见附录 A 的 A.4 节。

$$\left[\bigcap_{i=1}^{n} E_i\right]^c = \bigcup_{i=1}^{n} E_i^c \tag{1.4-8}$$

这个关系称为摩根(De Morgan)定律，这是由英国数学家摩根(1806－1871)发现的。

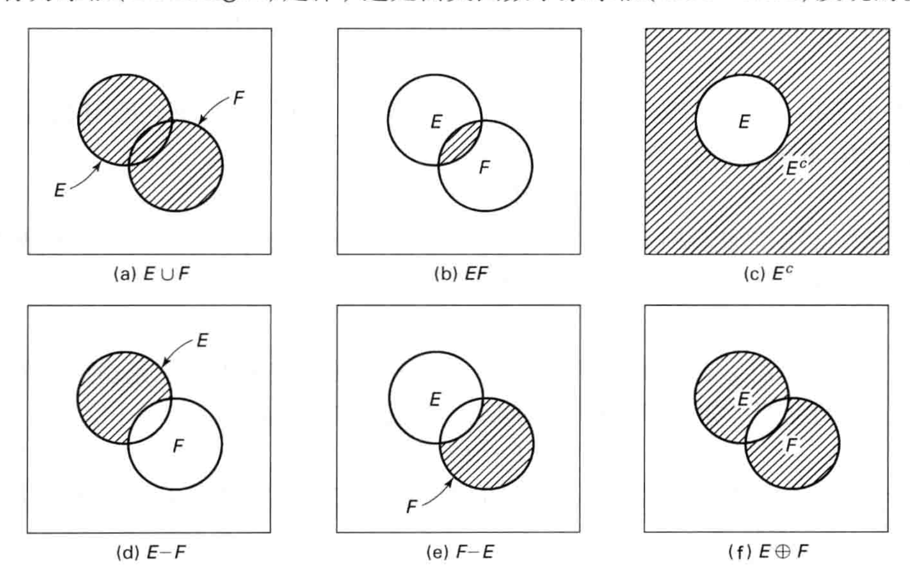


图 1.4-1 集合运算的维恩图

虽然维恩图使集合运算的结果可视化，但并不是真正地证明这些结果。为此，考虑两个集合相等的数学定义：如果集合 E 中的所有元素都在集合 F 中，反过来也是一样，则称 E 和 F 相等，或等价地描述为

$$E = F \text{ 如果 } E \subset F \quad \text{和} \quad F \subset E \tag{1.4-9}$$

例 1.4-7 (证明集合相等)如果要严格地证明两个集合相等，如式(1.4-4)的 $F = FE \cup FE^c$，那么证明过程如下：首先证明 $F \subset FE \cup FE^c$，然后证明 $F \supset FE \cup FE^c$。为了要证明 $F \subset FE \cup FE^c$，考虑任意元素 $\zeta \in F$，那么，对于任意的 E，ζ 也肯定在 FE 中或者在 FE^c 中，这样，ζ 肯定在 $FE \cup FE^c$ 中，这就证明了 $F \subset FE \cup FE^c$。另一方面，要证明 $F \supset FE \cup FE^c$，我们必须从任意元素 $\zeta \in FE \cup FE^c$ 开始，它肯定属于 FE 或 FE^c 中，所以，ζ 肯定属于 F 中，这就证明了 $F \supset FE \cup FE^c$。由于证明了两个集合互为包含，我们可以写成 $F = FE \cup FE^c$，即两个集合是相等的。

利用这样的方法，可以证明下列集合论中有用的定律：

1. 并的结合律

$$A \cup (B \cup C) = (A \cup B) \cup C$$

2. 交的结合律

$$A(BC) = (AB)C$$

3. 并的分配率

$$A \cup (BC) = (A \cup B)(A \cup C)$$

4. 交的分配率

$$A(B\cup C)=(AB)\cup(AC)$$

我们将使用这些恒等式或定律来分析下面的集合等式。然而，这四个定律必须先证明。这里，作为一个例子，我们给出定律 1 的证明。

例 1.4-8　(并的结合律的证明)我们希望证明 $A\cup(B\cup C)=(A\cup B)\cup C$。为此，必须证明 $A\cup(B\cup C)\subseteq(A\cup B)\cup C$ 和 $A\cup(B\cup C)\supseteq(A\cup B)\cup C$，先从前者开始，假定 $\zeta\in A\cup(B\cup C)$，那么可以得出 ζ 在 A 中或者在 $B\cup C$ 中，但是另一方面，ζ 在 A 中或在 B 中或在 C 中，所以 ζ 肯定在 $A\cup B$ 中或 C 中，这也等于在说 $\zeta\in(A\cup B)\cup C$。为了完成证明，我们走另一条路，从 $\zeta\in(A\cup B)\cup C$ 开始，那 ζ 肯定也是 $A\cup(B\cup C)$ 的元素，这一部分留给学生自己进行证明。

σ 域　考虑一个全集 Ω 和 Ω 的子集构成的集合簇，令 E 和 F 是这个集合簇中的任意两个子集，如果

(1) $\phi\in\mathscr{M}$、$\Omega\in\mathscr{M}$。

(2) 如果 $E\in\mathscr{M}$ 和 $F\in\mathscr{M}$，那么 $E\cup F\in\mathscr{M}$ 和 $EF\in\mathscr{M}$①。

(3) 如果 $E\in\mathscr{M}$，则 $E^c\in\mathscr{M}$。

那么这个集合簇形成了一个域 $\mathscr{M}$。

我们需要考虑集合的域(概率论中事件的域)，以便避免某些问题。如果事件的集合簇并不是一个域，那么可以为某些事件定义一个概率，但不能为它们的补集定义。即不能为没有发生的那些事件定义概率！类似地，我们需要考虑两个事件的并的概率，即两个事件的一个或两个发生的概率。因此，对于概率论来说，需要对域中的所有事件赋予一个概率。

我们常常要考虑实验结果和事件的无穷集，在这种情况下，必须扩展域的定义，一个 σ 域② $\mathscr{F}$ 也是一个域，它在任何可数的并、交和补都是闭的。因此，如果 E_1，⋯，E_n，⋯属于 $\mathscr{F}$，那么

$$\bigcup_{i=1}^{\infty}E_i \quad \text{和} \quad \bigcap_{i=1}^{\infty}E_i$$

也属于 $\mathscr{F}$，其中这两项简单定义为

$$\bigcup_{i=1}^{\infty}E_i \triangleq \{\text{所有处于至少一个 } E_i \text{ 的那些元素的集合}\}$$

和

$$\bigcap_{i=1}^{\infty}E_i \triangleq \{\text{处于每个 } E_i \text{ 的那些元素的集合}\}$$

注意，在没有特定定义的情况下，并和交的这两个无穷的运算是没有意义的。不像无穷求和是由数的极限来定义的，对集合没有这样的极限运算，因此需要给出一个定义。

事件　考虑一个样本空间为 Ω 的概率实验，如果 Ω 有可数个元素，那么，Ω 的每一个子集都可用一种与下一节给出的公理一致的方式赋予一个概率，然后，所有子集的类将组成一个域或者 σ 域，因为每一个子集都包括了。Ω 的所有子集的集合簇称为最大的 σ 域。

---

① 从这里由数学归纳法可以得出，如果 E_1，⋯，E_n 属于 $\mathscr{M}$，那么 $\cup_{i=1}^{n}E_i\in\mathscr{M}$ 和 $\cap_{i=1}^{N}E_i\in\mathscr{M}$。

② 有时也叫 σ 代数。

虽然有时我们并没有足够的概率信息为每一个子集赋予一个概率，在这种情况下，我们需要定义事件的一个小一点的域，这个域仍然是 σ 域，但它是一个较小的域，在下面的例子中我们讨论这一方法。在极限情况下，如果没有足够的概率信息，那么就必须用事件的最小 σ 域来满足我们的要求，这里的事件是由零事件 ϕ 和必然事件 Ω 组成，而这两个事件的集合簇是没有什么用处的 σ 域。

例 1.4-9 （由两个事件生成的域）假定在任意样本空间 Ω 只对两个事件 A 和 B 感兴趣，我们希望包含这两个事件的最小域。可以这样进行，首先产生一个样本空间的互斥的分解

$$\begin{aligned}\Omega &= \Omega(A \cup A^c)(B \cup B^c) \\ &= AB \cup AB^c \cup A^cB \cup A^cB^c\end{aligned}$$

然后从这四个基本的互斥事件(没有重叠)生成一个事件簇如下：首先，四个事件是 AB，AB^c，A^cB 和 A^cB^c，然后加上这些互斥事件的两个一组的并集：$AB \cup AB^c$，$AB \cup A^cB$ 和 $AB \cup A^cB^c$，最后再加上这四个互斥事件的三个一组的并集，那么，由于四个基本互斥事件的每一个可能包含或没有包含在事件中，总的事件数是 $2 \times 2 \times 2 \times 2 = 2^4 = 16$。

这个事件簇保证是一个域，因为我们是从来自于四个互斥事件的 16 个事件构成的。这样也保证了补集也是在集合簇中，这是通过 $\Omega = AB \cup AB^c \cup A^cB \cup A^cB^c$ 来保证的。并在集合簇中也是很平常的，因为在集合簇中的所有事件都是由四个基本互斥事件构成的，补集刚好是剩余的事件，如 $(AB \cup AB^c)^c = A^cB \cup A^cB^c$ 被认为是在集合簇中。因此我们有一个域，事实上这是一个最小的域，它只包含 A 和 B，称为由 A 和 B 产生的域。你能证明事件 A 也在这个域中吗?

当 Ω 不是可数的，例如，当 $\Omega = R^1$ = 实数直线(real line)，近代数学(测度论)发现，并非 Ω 的每一个子集都能以一致的方式赋予概率(可测的)，所以我们必须用整个事件 Ω 的子集的较小一点的集合簇来满足我们的要求，这个较小的集合簇形成了 σ 域。例如，在实数直线 R^1 上，我们能够从所有的开或闭区间产生一个 σ 域，称为实数直线上事件的博雷尔(Borel)域。实际上，实数直线上的博雷尔域包括了所有感兴趣的工程和科学中的所有子集①。

到目前为止，我们已经有了概率的公理化理论所要求的三个目标中的两个，这两个称为实验结果 ζ 的样本空间和定义在样本空间 Ω 上事件的 σ 域 $\mathscr{F}$，我们还需要一个概率测度 P，三个目标 $(\Omega, \mathscr{F}, P)$ 形成了三元组，称为概率空间 $\mathscr{P}$，它构造了我们的数学模型。然而，概率测度 P 必须满足由科尔莫戈罗夫提出的下面三个公理。

1.5 概率的公理化定义

概率是一个集合函数 $P[\cdot]$，它对每个事件 $E \in \mathscr{F}$ 指定一个数 $P[E]$，这个数称为事件 E 的概率，它满足三个条件：

(1) $P[E] \geqslant 0$。 (1.5-1)

(2) $P[\Omega] = 1$。 (1.5-2)

(3) $P[E \cup F] = P[E] + P[F]$ 如果 $EF = \phi$。 (1.5-3)

---

① 对于二维欧几里得样本空间，事件的博雷尔域是 $R^1 \times R^1 = R^2$ 的子集，对于三维样本空间，博雷尔域应是 $R^1 \times R^1 \times R^1 = R^3$ 的子集。

概率测度不像一个普通函数，它并不是对一个变量而是对一个集合指定一个数，即它是集合的测度，这是事件的数学模型。由于这是一个特殊函数，为了区分它，我们总是对它的自变量使用方括号，这里自变量是在样本空间 Ω 中的实验结果 ζ 的集合。

上面三个公理足以建立下列基本结论①，除了一个结论，其他留给读者作为习题。令 E 和 F 是包含在 $\mathscr{F}$ 中的事件，那么

(4) $P[\phi]=0$。 (1.5-4)

(5) $P[EF^c]=P[E]-P[EF]$。 (1.5-5)

(6) $P[E]=1-P[E^c]$。 (1.5-6)

(7) $P[E\cup F]=P[E]+P[F]-P[EF]$。 (1.5-7)

由公理 3，用数学归纳法可以证明如下关系：

$$P\left[\bigcup_{i=1}^{n} E_i\right]=\sum_{i=1}^{n} P[E_i] \text{ 如果 } E_iE_j=\phi \text{ 对所有 } i\neq j \tag{1.5-8}$$

根据这个结果和式(1.5-7)，我们能够证明 $P[\cup_{i=1}^{n}E_i]\leqslant\sum_{i=1}^{n}P[E_i]$。这个结果有时也称为并的边界，常常用在数字通信理论中提供误差概率的上限。

例 1.5-1　(两事件并的概率)我们按结论 7 进行证明。首先将事件 $E\cup F$ 分解为三个互斥事件，

$$E\cup F=EF^c\cup E^cF\cup EF$$

由公理 3

$$\begin{aligned}P[E\cup F]&=P[EF^c\cup E^cF]+P[EF]\\&=P[EF^c]+P[E^cF]+P[EF],\text{ 再由公理 3}\\&=P[E]-P[EF]+P[F]-P[EF]+P[EF]\\&=P[E]+P[F]-P[EF]\end{aligned} \tag{1.5-9}$$

我们可以将这一结论应用到下面的问题中。

在某个面包店中，两个感兴趣的事件是 $W\triangleq\{$提供白面包$\}$和 $R\triangleq\{$提供黑麦面包$\}$，根据以往经验，取 $P[W]=0.8$，$P[R]=0.7$，我们也知道两种面包同时提供的概率为 0.6，即 $P[WR]=0.6$，两种面包至少有一种提供的概率是多少？即 $P[W\cup R]$是多少呢？由结论 7 得到的答案为

$$\begin{aligned}P[W\cup R]&=P[W]+P[R]-P[WR]\\&=0.8+0.7-0.6\\&=0.9\end{aligned}$$

我们为一个术语暂停一下，当实验结果是 E 中的元素时，就说事件 E 发生了，所以，“$P[E]$”读成为事件 E 发生的概率。

事件的测度，而非实验结果的测度　读者可能注意到我们谈论的是事件的概率而不是实验

---

① 第四个公理：对于所有 $i\neq j$，$E_iE_j=\phi$，则 $P[\cup_{i=1}^{\infty}E_i]=\sum_{i=1}^{\infty}P[E_i]$，如果按严格处理极限和可数并，必须包括这一公理，这里我们暂且不涉及它，后面的章节再介绍。

结果的概率，对于有限和可数的样本空间，我们谈论的刚好是实验结果的概率，然而不这样做的原因有几个，其一是即使谈论实验结果的概率，我们也仍然需要事件的概率，这样就需要两种类型的概率测度，一个为实验结果，一个为事件；其二，有时我们只知道某些事件的概率，而并不清楚对每个实验结果赋概率值的细节；最后，也是最重要的是，对于连续的样本空间，实验结果是不可数的，例如，在实数区间[0，5] 中的点，它们具有零概率，因此，基于实验结果的概率的理论是没有用的，所以，我们采用事件的概率测度，而不是采用实验结果的概率测度。

例 1.5-2 （投掷硬币一次）实验由投掷硬币一次组成，实验结果是 H 和 T，样本空间为

$$\Omega=\{\mathrm{H},\mathrm{T}\}$$

事件的 σ 域由下列集合构成：{H}，{T}，Ω，ϕ，假定硬币是均匀的，则①

$$P[\{\mathrm{H}\}]=P[\{\mathrm{T}\}]=\tfrac{1}{2},\qquad P[\Omega]=1,\qquad P[\phi]=0$$

例 1.5-3 （投掷骰子一次）实验由投掷骰子一次组成，实验结果是点数 $\zeta=1,\cdots,6$，样本空间为 $\Omega=\{1,2,3,4,5,6\}$ 事件域由 2^6 个事件组成，每个事件包含或没有包含实验结果 i，某些事件是

$$\phi,\Omega,\{1\},\{1,2\},\{1,2,3\},\{1,4,6\}\ \ \text{和}\ \ \{1,2,4,5\}$$

我们为每个基本事件或单一事件 $\{\zeta\}$ 赋予一个概率

$$P[\{\zeta\}]=\tfrac{1}{6},\qquad i=1,\cdots,6$$

所有概率就可以依据基本公理和对基本事件假定的概率进行计算。例如，假定 $A=\{1\}$，$B=\{2,3\}$，则 $P[A]=\dfrac{1}{6}$；又由于 $AB=\phi$，所以，$P[A\cup B]=P[A]+P[B]$；此外，$P[B]=P[\{2\}+\{3\}]=\dfrac{2}{6}$，所以

$$P[A\cup B]=\tfrac{1}{6}+\tfrac{2}{6}=\tfrac{1}{2}$$

例 1.5-4 （从罐中选择球）假定从一个盛有 12 个球的罐子中随机捡出一个球，每个球分别标有数字 1～12，实验结果理想化为数 $\zeta=1$ 到 12，样本空间为

$$\Omega=\{1,\cdots,12\}$$

令下列事件被指定：

$$A^{②}=\{1,\cdots,6\},\qquad B=\{3,\cdots,9\}$$
$$A\cup B=\{1,\cdots,9\},\qquad AB=\{3,4,5,6\},\qquad AB^c=\{1,2\}$$
$$B^c=\{1,2,10,11,12\},\qquad A^c=\{7,\cdots,12\},\qquad A^cB^c=\{10,11,12\}$$
$$(AB)^c=\{1,2,7,8,9,10,11,12\}$$

因此

$$P[A]=P[\{1\}]+P[\{2\}]+\cdots+P[\{6\}]$$
$$P[B]=P[\{3\}]+\cdots+P[\{9\}]$$
$$P[AB]=P[\{3\}]+\cdots+P[\{6\}]$$

如果 $P[\{1\}]=\cdots=P[\{12\}]=\dfrac{1}{12}$，则 $P[A]=\dfrac{1}{2}$，$P[B]=\dfrac{7}{12}$，$P[AB]=\dfrac{4}{12}$，等等。

---

① 记住，实验结果(outcome)ζ 是我们实验的输出(output)或结果(result)。

② 当我们从罐中取出的球为 1～6 时，我们说事件 A 发生。

我们要指出的是概率论也可以从稍微不同的公理集[1 ~7]建立起来，然而，无论采用什么公理，无论建立什么样的理论，都必须把概率的经验概念用相对频率来建模，因为它在解决实际问题时是很有用的。

事件并的概率　将式(1.5-7)扩展到三个事件的情况是相当简单的，但有点冗长乏味。我们考虑三个事件 E_1，E_2，E_3，希望计算这三个事件至少有一个发生的概率 $P[\cup_{i=1}^{3}E_i]$。由维恩图1.5-1，我们看到在 $\cup_{i=1}^{3}E_i$ 中的7个互斥的区域，记为 Δ_i，$i=1$，…，7，可以采用例1.4-9同样的方法证明。由公理3可得，$P[\cup_{i=1}^{3}E_i] = P[\cup_{i=1}^{7}\Delta_i] = \sum_{i=1}^{7}P[\Delta_i]$。

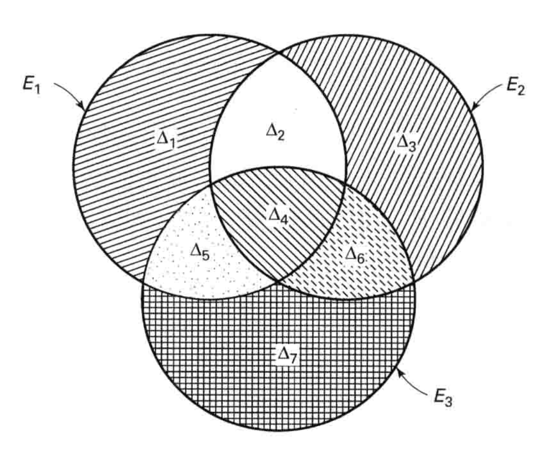


图1.5-1　将 $\cup_{i=1}^{3}E_i$ 划分为七个互斥的区域Δ_1，…，Δ_7

根据原始的事件，7个互斥的区域可以标识为

$$\Delta_1 = E_1E_2^cE_3^c = E_1(E_2 \cup E_3)^c$$

$$\Delta_2 = E_1E_2E_3^c$$

$$\Delta_3 = E_1^cE_2E_3^c = E_2(E_1 \cup E_3)^c$$

$$\Delta_4 = E_1E_2E_3$$

$$\Delta_5 = E_1E_2^cE_3$$

$$\Delta_6 = E_1^cE_2E_3$$

$$\Delta_7 = E_1^cE_2^cE_3 = E_3(E_1 \cup E_2)^c$$

概率 $P[\Delta_i]$($i=1$，…，7) 的计算由式(1.5-5)和式(1.5-7)得出，因此，可以计算得出

$$P[\Delta_1] = P[E_1] - P[E_1E_2 \cup E_1E_3]$$
$$= P[E_1] - \{P[E_1E_2] + P[E_1E_3] - P[E_1E_2E_3]\}$$

上式的第一行用式(1.5-5)得到，而第二行由式(1.5-7)得到。而 $P[\Delta_i]$($i=3$，7) 的计算与 $P[\Delta_1]$ 的计算类似，采用相同的步骤，得

$$P[\Delta_3] = P[E_2] - \{P[E_1E_2] + P[E_2E_3] - P[E_1E_2E_3]\}$$

$$P[\Delta_7] = P[E_3] - \{P[E_1E_3] + P[E_2E_3] - P[E_1E_2E_3]\}$$

而 $P[\Delta_2]$，$P[\Delta_5]$ 和 $P[\Delta_6]$ 的计算也非常类似，其中要应用式(1.5-5)，于是

$$P[\Delta_2] = P[E_1E_2] - P[E_1E_2E_3]$$

$$P[\Delta_5] = P[E_1E_3] - P[E_1E_2E_3]$$

$$P[\Delta_6] = P[E_2E_3] - P[E_1E_2E_3]$$

最后

$$P[\Delta_4] = P[E_1E_2E_3]$$

回想一下 $P[\cup_{i=1}^{i=3}E_i] = \sum_{i=1}^{7}P[\Delta_i]$，我们只需要将所有 $P[\Delta_i]$ 加起来就可以得到期望的结果，由此可得

$$P\left[\bigcup_{i=1}^{3}E_i\right]=\sum_{i=1}^{3}P[E_i]-\Big(P[E_1E_2]+P[E_1E_3]+P[E_2E_3]\Big)+P[E_1E_2E_3] \quad (1.5\text{-}10)$$

注意，这个结果是可以理解，因为在加三个事件的测度过程中，我们计数了双重叠区域两次，但如果减去这三个重叠区域，那我们又根本没有计数 $E_1E_2E_3$，所以还必须把它加回来。如果采用符号 $P_i \triangleq P[E_i]$，$P_{ij} \triangleq P[E_iE_j]$ 和 $P_{ijk} \triangleq P[E_iE_jE_k]$，其中 $1 \leqslant i < j < k \leqslant 3$，重写式(1.5-10)为

$$P\left[\bigcup_{i=1}^{3}E_i\right]=\sum_{i=1}^{3}P_i-\sum_{1\leqslant i<j\leqslant 3}P_{ij}+\sum_{1\leqslant i<j<k\leqslant 3}P_{ijk}$$

最后一个和式只包含一项，称为 P_{123}。用符号 S_l 来表示每个和式，其中 l 表示和式项数的下标，那么

$$P\left[\bigcup_{i=1}^{3}E_i\right]=S_1-S_2+S_3,\quad 其中 S_1\triangleq\sum_{i=1}^{3}P_i,\ S_2\triangleq\sum_{1\leqslant i<j\leqslant 3}P_{ij} 和$$

$$S_3\triangleq\sum_{1\leqslant i<j<k\leqslant 3}P_{ijk}$$

为什么要引入一个新的符号呢？利用符号 S_l，$l=1$，…，我们可以把式(1.5-10)的结果扩展到一般情况。

定理1.5-1　(n 个事件并的概率)在一给定的实验中，在事件 $E_1, E_2, \cdots, E_n$ 中至少有一个事件发生的概率 P 为

$$P=S_1-S_2+\cdots\pm S_n$$

其中，$S_1 \triangleq \sum_{i=1}^{n}P_i$，$S_2 \triangleq \sum_{1\leqslant i<j\leqslant n}^{n}P_{ij}$，…，$S_n \triangleq \sum_{1\leqslant i<j<k<\cdots<l\leqslant n}P_{ijk\cdots l}$ 最后一个和式有 n 个下标，并只有一项。

这个定理的证明在[1 ~8, p.89]中给出，也可以用数学归纳法证明，即假定 $P=S_1-S_2+\cdots\pm S_n$ 是正确的，然后只要证明对于 $n+1$，$P=S_1-S_2+\cdots\mp S_{n+1}$ 成立，我们把这个留给勇敢的读者去证明。

1.6　联合概率、条件概率、全概率和独立性

假定我们进行下面的实验：在美国某个城市，希望收集该城市的气象数据，我们特别对三个事件感兴趣，称为 A，B 和 C，其中

A 是在任意特定的一天，温度等于或超过 10℃的事件。

B 是在任意特定的一天，降雨量等于或超过 5 mm 的事件。

C 是在任意特定的一天，A 和 B 同时出现的事件，即 $C \triangleq AB$。

由于 C 是一个事件，可以计算 $P[C]=P[AB]$，称 $P[AB]$ 是事件 A 和 B 的联合概率，很显然，这一表示可以扩展到超过两个事件的情形，如 $P[EFG]$ 是事件 E，F 和 G① 的联合概率。

① E，F，G 是定义在同一个概率空间上的三个事件。

令 n_E表示事件 E 发生的天数。在 1000 天内($n = 1000$)，得到下面的观测，$n_A = 811$，$n_B = 306$，$n_{AB} = 283$。由概率的相对频率来解释

$$P[A] \approx \frac{n_A}{n} = \frac{811}{1000} = 0.811$$

$$P[B] \approx \frac{n_B}{n} = 0.306$$

$$P[AB] \approx \frac{n_{AB}}{n} = 0.283$$

现在考虑一下比值 n_{AB}/n_A，这是在给定事件 A 发生的条件下事件 AB 发生的相对频率，应用到这里就是在给定温度等于或超过 10℃时降水量等于或超过 5 mm 的时间比值，因此，我们处理的是在另一个事件给定或者另一个事件已经发生的条件下事件的频率。注意：

$$\frac{n_{AB}}{n_A} = \frac{n_{AB}/n}{n_A/n} \approx \frac{P[AB]}{P[A]} \tag{1.6-1}$$

上面依据经验的概念给我们引入了条件概率测度。

条件概率　条件概率 $P[B|A]$定义为

$$P[B|A] \triangleq \frac{P[AB]}{P[A]}, \quad \text{如果 } P[A] > 0 \tag{1.6-2}$$

读成“给定事件 A 已经发生的条件下事件 B 发生的概率”。类似地，我们有

$$P[A|B] \triangleq \frac{P[AB]}{P[B]}, \quad \text{如果 } P[B] > 0 \tag{1.6-3}$$

定义式(1.6-2)和式(1.6-3)能够用来计算 AB 的联合概率，因为

$$P[AB] = P[A|B]P[B]$$
$$= P[B|A]P[A]$$

独立性

定义(事件的独立性)　(i)两个事件 $A \in \mathscr{F}$, $B \in \mathscr{F}$，且 $P[A] > 0$，$P[B] > 0$，当且仅当

$$P[AB] = P[A]P[B] \tag{1.6-4}$$

时，称事件 A 和 B 是独立的。

一般来说，由于 $P[AB] = P[B|A]P[A] = P[A|B]P[B]$，因此，对于独立事件，有

$$P[A|B] = P[A] \tag{1.6-5a}$$

$$P[B|A] = P[B] \tag{1.6-5b}$$

因此，这一定义满足了我们的直观理解，如果 A 和 B 是独立的，那么 B 的实验结果对 A 的条件概率没有影响，反过来也一样。

(ii) 有三个定义在$\mathscr{F}$上的事件 A，B，C，并且具有非零概率，当且仅当

$$P[ABC] = P[A]P[B]P[C] \tag{1.6-6a}$$

$$P[AB] = P[A]P[B] \tag{1.6-6b}$$

$$P[AC] = P[A]P[C] \tag{1.6-6c}$$

$$P[BC] = P[B]P[C] \tag{1.6-6d}$$

称它们是联合独立的。这是(i)的扩展，对 n 个事件 A_1，…，A_n 的独立性的定义也可以采用类似的方法。注意，$P[ABC]=P[A]P[B]P[C]$ 不是充分条件，两两独立性还必须证明。

(iii)令 $A_i(i=1, \cdots, n)$ 是包含在 $\mathscr{F}$ 中的事件，当且仅当

$$P[A_iA_j]=P[A_i]P[A_j]$$

$$P[A_iA_jA_k]=P[A_i]P[A_j]P[A_k]$$

$$\vdots$$

$$P[A_1\cdots A_n]=P[A_1]P[A_2]\cdots P[A_n]$$

上式对所有下标的组合 $1\leqslant i<j<k<\cdots\leqslant n$ 都成立，那么，我们说 $\{A_i\}$ 是联合独立。

例 1.6-1 (骰宝，Sic bo)骰宝是一种赌场玩的赌博游戏，玩者对三个骰子的投掷的结果下注，不同的赌注有不同的赔率，下面列出两个有关的赔率：

(1)指定三同号(180∶1)，即由打赌者预先指定

(2)没有指定的三同号(30∶1)，即任意三同号。

从打赌者的观点看，他赢的概率是多少呢，它期待赢取的又是多少呢？

解

(1)(指定三同号)令 E_i 表示在第 i 次投掷中出现的指定的实验结果，那么，三个号 $E_1E_2E_3$ 同时出现的概率 $P[E_1E_2E_3]=P[E_1]P[E_2]P[E_3]=1/216$，这里利用了这样一个事实，三个事件是独立的，因为它们是三次不同的投掷。公平的赔率是 216∶1，而不是 180∶1。

(2)(没有指定的三同号)第一次投掷可以出现任意数字，而后两次投掷必须匹配第一次投掷的数字，因此，P[没有指定的三同号]$=1\times1/6\times1/6=1/36$，因此公平的赔率是 36∶1，而不是 30∶1。

例 1.6-2 (检验三个事件的独立性)一个罐子中装有 10 标有数字的黑色球(有些是偶数，有些是奇数)和 20 个标有数字的白色球(有些是偶数，有些是奇数)，在每一种颜色的球中，有一些球比其他一些球要轻，罐子的精确分解如图 1.6-1 所示。实验结果是三元变量 $\boldsymbol{\zeta}=$(颜色，质量，数字)，样本空间 Ω 是所有这些三元变量的集合。每次抽取是完全随机的。

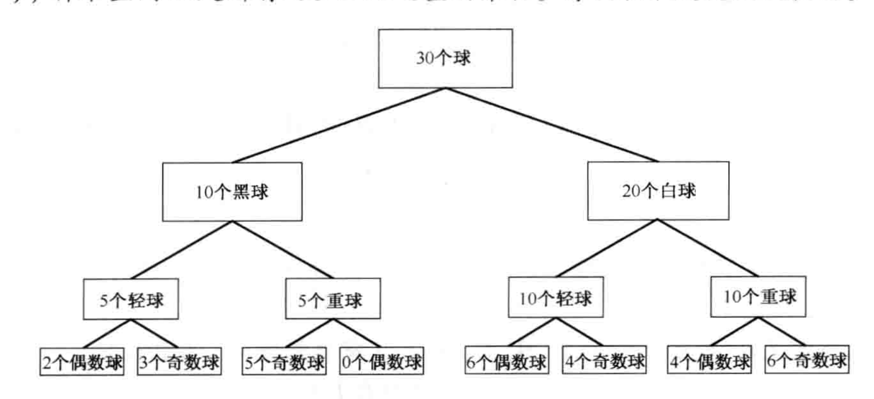


图 1.6-1 罐子的分解图

令 A 表示抽到一个黑球的事件，B 表示抽到一个轻球的事件，C 表示抽到一个偶数球的事件，A，B，C 是独立的吗？

解 我们首先检验是否 $P[ABC]=P[A]P[B]P[C]$，由于有 1/3 的球是黑色的，所以，

$P[A]=1/3$。又由树图可以看出，球的 15/30 是轻球，所以 $P[B]=1/2$，球的 12/30 是偶数，所以 $P[C]=2/5$。由于事件 ABC 是黑色、轻的和偶数的球且只有两个，那么，$P[ABC]=2/30$。而 $P[ABC]=P[A]P[B]P[C]$ 直接相乘得出的结果也是 2/30，因此，三个事件通过独立性的这一部分的测试，然而，对于独立而言，我们还必须有 $P[AB]=P[A]P[B]$，$P[AC]=P[A]P[C]$ 和 $P[BC]=P[B]P[C]$，注意到 $P[AC]=2/30$，而 $P[A]P[C]=1/3\times12/30=2/15\neq2/30$，因此，$A$，$B$ 和 C 不是联合独立的。

复合事件

我们经常需要考虑复合实验或重复试验。如果有一个定义在每一个实验上的概率空间，那么就会明白整体的或复合的实验是什么。有两种情况需要考虑，一种情况就是对物理事实建模常常需要重复实验，这些重复实验看起来是相互独立的；另一种重要的情况就是实验结果似乎依赖于早期试验的结果。

独立实验　考虑两个独立实验，也就是说一个实验结果不受另一个的过去、现在和将来的实验结果影响。令每个实验都有它自己的样本空间 Ω、实验结果 ζ 和概率测度 P。特别是我们有

$$\zeta_1\in E_1\subset\Omega_1 \text{概率测度为} P_1 \quad \text{和}$$
$$\zeta_2\in E_2\subset\Omega_2 \text{概率测度为} P_2$$

如图 1.6-2 所示。

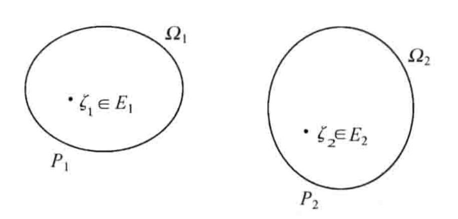


图 1.6-2　两个复合概率实验

我们希望能够与复合实验联系起来，即复合实验的样本空间是两个样本空间的笛卡儿积，

$$\Omega\triangleq\Omega_1\times\Omega_2$$

具有向量的实验结果(元素)$\zeta=(\zeta_1,\zeta_2)\in E\subset\Omega$。

例 1.6-3　(投掷两枚硬币)假定有两个实验，每一个实验是由投掷一枚双面硬币组成的，两面分别用 H 和 T 表示，那么，有 $\Omega_1=\{\mathrm{H},\mathrm{T}\}=\Omega_2$，在复合实验中，有 $\Omega=\{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\}$，我们也正好用一串字符 H 和 T 而不是向量写出实验结果 $\zeta\in\Omega$，用这样的表示，我们有 $\Omega=\{\mathrm{HH},\mathrm{HT},\mathrm{TH},\mathrm{TT}\}$。在第一个实验中，考虑事件 $E_1=\{\mathrm{T}\}$，在第二个实验中 $E_2=\{\mathrm{H}\}$，那么，在复合实验中，有事件 $E=\{\mathrm{TH}\}=E_1\times E_2\subset\Omega$。

当用互积表示集合时，我们有

$$E_1\times E_2\triangleq\{\zeta=(\zeta_1,\zeta_1)|\zeta_1\in E_1 \text{ 和 } \zeta_2\in E_2\}$$

所以，在两个集合的互积中的元素是所有可能的元素对，元素对的每一个元素来自于每一个集合。

例 1.6-4　(投掷骰子 2 次)假定有两个实验，每一个实验是由投掷一枚骰子组成，骰子有六个面，分别用实验结果 1 ~6 表示，那么，有 $\Omega_1=\{1,2,3,4,5,6\}=\Omega_2$，在复合实验中，把两个样本空间的互积的一对元素作为实验结果，即 $\Omega=\{11,12,\cdots,16,21,\cdots,26,\cdots,61,\cdots,66\}=\Omega_1\times\Omega_2$，注意，所有事件($\Omega$ 的子集)不是 $E_1\times E_2$ 的形式，事实上，这是一种特殊情况。例如，考虑事件 $\{11,12,31\}$，它是 $\{1,3\}\times\{1,2\}$ 中不含实验结果 32 的那个事件。然而，可以把这个事件写成集合的互积的互不相交的集合的并

$$\{11,12,31\}=\{\{1\}\times\{1,2\}\}\cup\{\{3\}\times\{1\}\}$$

我们常常对物理实验中那些相互独立的不相交的模型感兴趣，这要求一个定义。因此，对于复合事件，如果事件 E 的概率能够根据各自的概率测度 P_1 和 P_2 来表示，那么就定义这两个复合事件在数学上是独立的。

定义 1.6-1 对于两个实验，如果：

(i) 对于互积事件 $E=E_1\times E_2$，我们能够写成

$$P[E]\triangleq P_1[E_1]P_2[E_2]$$

(ii) 在复合实验中，一般事件 E 的概率能够根据单一事件的概率来表示

$$P[E]\triangleq\sum_{(\zeta_1,\zeta_2)\in E}P_1[\{\zeta_1\}]P_2[\{\zeta_2\}]$$

那么，称这两个实验是独立的。

我们可以把这个概念推广到 n 个实验的组合，得到的复合实验的样本空间为

$$\begin{aligned}\Omega&\triangleq\bigotimes_{i=1}^{n}\Omega_i\\&=\Omega_1\times\Omega_2\times\Omega_3\times\cdots\times\Omega_n\end{aligned}$$

向量(字符串)的实验结果 $\zeta=(\zeta_1,\cdots,\zeta_n)\in E\subset\Omega$，$\Omega$ 是复合实验的样本空间。

例 1.6-5 (三个实验)考虑三个独立的实验，每一个都有自己的样本空间 Ω_i，$i=1,2,3$，令 E_i 是样本空间 Ω_i 中的任意事件，那么，在复合实验中的一般互积事件的概率为

$$P[E_1\times E_2\times E_3]=P_1[E_1]P_2[E_2]P_3[E_3]$$

其中，事件 E_i 是由 Ω_i 中的可测的子集的并和交组成。

例 1.6-6 (重复的硬币投掷实验)考虑投掷一枚硬币 n 次，每次投掷都认为是一个随机的、独立的实验，令在每次实验中单个实验结果用 H 和 T 表示，那么在复合实验中的实验结果是由 H 和 T 组成、长度为 n 的字符串，总共有 2^n 个不同的有序字符串。字符串有 kH 和 $(n-k)T$ 的概率为

$$\begin{aligned}P[(\zeta_1,\cdots,\zeta_n)]&=\prod_{i=1}^{n}P_i[\{\zeta_i\}]\\&=p^kq^{n-k}\end{aligned}$$

其中 p 和 $q\triangleq1-p$ 是单次硬币投掷出现 H 和 T 的概率，且 $0\leqslant p\leqslant1$。

我们也可以根据一般事件而不单独的事件来表达这些复合概率。考虑两个复合事件，概率空间为

$$\zeta_1\in E_1\subset\Omega_1\text{ 且概率测度为 }P_1,\ \zeta_2\in E_2\subset\Omega_2\text{，且概率测度为 }P_2$$

那么，复合实验由下列概率空间组成：

$$\zeta\triangleq(\zeta_1,\zeta_2)\in E\subset\Omega\text{ 且概率测度为 }P$$

其中复合概率测度为 P，它是为事件 $E\subset\Omega$ 定义的，其定义如下：首先，必须将 E 中的复合事件写成来自两个实验的互积事件的互不相交事件的并集，即

$$E=\bigcup_{i=1}^{k}E_{1,i}\times E_{2,i}$$

其中 k 为正整数，$E_{1,i}$ 和 $E_{2,i}$ 分别是 Ω_1 和 Ω_2 中的事件，最简单的情况是 E 自身是一个互积事件，即 $k=1$，但是，正如例 1.6-4 所看到的那样，通常必须取几个互积事件的并来表达复合空间中的任意事件 E。

定义 1.6-2　（另一种定义）我们说事件是独立的，指的是对于复合实验中的任意事件 E，

$$P[E] \triangleq \sum_{i=1}^{k} P_1[E_{1,i}]P_2[E_{2,i}],\quad \text{其中 } E = \bigcup_{i=1}^{k} E_{1,i} \times E_{2,i}$$

是一个互不相交事件的并集，$E_{1,i}$和 $E_{2,i}$分别是 Ω_1 和 Ω_2 中的事件，这里 k 表达复合事件 E 所必须的互积事件数。

注意，由于事件是互不相交的，所以概率的相加性是适合的，我们立即看到这个定义与前面根据单一事件或基本事件给出的定义是一致的。只要取 $E_{1,i}$和 $E_{2,i}$为单一事件就很容易看出这一点。很显然，这一更为通用的方法也可以扩展到 $n(n>2)$ 个实验的情况。下面我们转到更为复杂的多个独立实验的情况。

非独立实验[*①]　考虑两个"非独立的实验"，这意味着第二个实验的概率将依赖于第一实验发生的事件。假定第一个实验由实验结果 $\zeta_{1,i}$，$(i = 1, 2, \cdots, k)$ 组成，它们的概率 $P_1[\{\zeta_{1,i}\}]$ 是给定的，第二个实验的概率测度肯定与来自于第一个实验的索引 i 有关，即

$$P_{2,i}[E_2],\ \text{对每个事件}\ E_2 \subset \Omega_2$$

其中 Ω_2 是第二个实验的样本空间，如图 1.6-3 所示。

图 1.6-3　两个非独立复合实验

那么，我们就能写出复合实验的概率测度如下所述。

定义 1.6-3　（非独立实验）令 $E_1 = \{\zeta_{1,i}\}$ 是样本空间 Ω_1 中第 i 个的单一事件，E_2是 Ω_2 中的事件，考虑复合实验中的互积实验 $E = E_1 \times E_2$，则

$$P[E] \triangleq P_1[\{\zeta_{1,i}\}]P_{2,i}[E_2]$$

其中第二个实验中的概率测度是第一个实验的实验结果的函数。

首先，我们注意到这个定义与前面的独立实验的定义是一致的，这是因为在独立事件的情况下，所有的 $P_{2,i}$是相同的，即对所有的 i，$P_{2,i} = P_2$。

更一般的情况，令第一个实验中的事件是 $E_1 = \cup_i \{\zeta_{1,i}\}$，即 i 个基本（单一）事件的并，那么，复合事件 $E = E_1 \times E_2$ 的概率可写为

$$P[E] \triangleq \sum_{i} P_1[\{\zeta_{1,i}\}]P_{2,i}[E_2]$$

这里的相加是因为，在第一次实验中，i 个基本事件$\{\zeta_{1,i}\}$中只有一个发生。

例 1.6-7　（投掷有偏差的硬币）考虑三枚有偏差的硬币，我们以概率 $p_1 = P_1[\{\mathrm{H}\}]$投掷第一枚硬币，根据实验结果是 H 还是 T，再分别投掷第二枚和第三枚硬币。假定第二枚硬币的概率为 $p_2 = P_2[\{\mathrm{H}\}]$，第三枚硬币的概率为 $p_3 = P_3[\{\mathrm{H}\}]$，当然，在这里假定所有 p_i满足 $0 < p_i < 1$。那么，对于 $p_2 \neq p_3$，我们有非独立实验的情况。例如，计算 $P\{\mathrm{HT}\}$，我们得到 $p_1(1-p_2)$，计算 $P\{\mathrm{TH}\}$，可得到$(1-p_1)p_3$，以此类推。

例 1.6-8　（以事件为条件的概率）考虑今天的天气分别为晴天、阴天和雨天，概率分别为 $p_{1,s}$，$p_{1,c}$和 $p_{1,r}$，三个概率之和为 1，那么明天天气也可能是晴天、阴天和雨天，这可能依

① 标有星号（*）的材料在第一次阅读时可以略去。

赖于今天的天气，所以，明天天气的条件概率依赖于三个条件事件，对每一个条件事件都有一个不同的测度，这样，对于第2天，我们就有一个由三个条件概率测度组成的集合，对第1天的每一个条件都有一个条件概率测度。

与条件概率的关系 考虑一个复合实验$(\Omega, \mathscr{F}, P)$，它是由两个复合实验$(\Omega_1, \mathscr{F}_1, P_1)$和$(\Omega_2, \mathscr{F}_2, P_2)$组成，这两个复合实验是相互独立的，所以，对于任意互积事件$E_1 \times E_2 \in \mathcal{F}$有

$$P[E_1 \times E_2] = P_1[E_1]P_2[E_2]$$

其中$E_1 \in \mathscr{F}_1$和$E_2 \in \mathscr{F}_2$。我们可以认为第一个实验是在第二个实验之前发生。令条件事件$B \in \mathscr{F}$具有$B = B_1 \times \Omega_2$的形式，其中$B_1 \in \mathscr{F}_1$，那么$P[B] = P_1[B_1] \cdot 1$。类似地，令$A \in \mathscr{F}$具有$A = \Omega_1 \times A_2$，其中$A_2 \in \mathscr{F}_2$，那么$P[A] = 1 \cdot P_2[A_2]$。于是，我们求得条件概率为

$$\begin{aligned}
P[A|B] &= \frac{P[AB]}{P[B]} \\
&= \frac{P[(\Omega_1 \times A_2) \cap (B_1 \times \Omega_2)]}{P_1[B_1]} \\
&= \frac{P[B_1 \times A_2]}{P_1[B_1]} \\
&= \frac{P_1[B_1]P_2[A_2]}{P_1[B_1]} \\
&= P_2[A_2]
\end{aligned}$$

其中注意到

$$\begin{aligned}
(\Omega_1 \times A_2) \cap (B_1 \times \Omega_2) &= \{(\zeta_1,\zeta_2)|\zeta_2 \in A_2\} \cap \{(\zeta_1,\zeta_2)|\zeta_1 \in B_1\} \\
&= B_1 \times A_2
\end{aligned}$$

对于两个独立的实验来说，这是我们期待的结果，但是，对于两个非独立的实验又会是什么结果呢？

***例 1.6-9** (非独立的情况)考虑如上例的复合实验，即$B = B_1 \times \Omega_2$和$A = \Omega_1 \times A_2$，但现在假定这些实验不是独立的，假定在第一个实验中的实验结果是一个有限数k，写出第二个实验的概率测度，这个测度是以第一个实验的实验结果作为函数的，即对于每一个实验结果$\zeta_{1,i} \in \Omega_1$，$i = 1, \cdots, k$，写出$P_{2,i}$的表达式。再假定$B_1 = \{\zeta_{1,i}\}$，那么，继续前一个例子，有

$$\begin{aligned}
P[A|B] &= \frac{P[(\Omega_1 \times A_2) \cap (B_1 \times \Omega_2)]}{P_1[B_1]} \\
&= \frac{P_1[B_1]P_{2,i}[A_2]}{P_1[B_1]} \\
&= P_{2,i}[A_2]
\end{aligned}$$

这是我们所期待的。

例 1.6-10 (通信信道与信源)在二元通信系统中，我们有一个二元信源S与二元信道C(如图1.6-4所示)，图中的标注是根据条件概率标注的。这个组合实验的样本空间Ω为$\Omega = \{\zeta = (x, y): x = 和 y = 0 或 1\} = \{(0,0), (0,1), (1,0), (1,1)\}$，其中，$x$表示信源的输出，它也是信道的输入，$y$表示信道的输出。那么，联合概率函数为$P[\{(x,y)\}] = P_S[\{x\}]P_C[\{y\}|\{x\}]$，$x,y = 0, 1$，其中$P_S$是信源$S$的概率测度，$P_C$是信道$C$的条件概率测度。

由于有噪声，发射 0 有时也有可能解码后被接收为 1，反之亦然。重复使用信道，得

$$P_C[\{0\}|\{0\}]=0.9,\qquad P_C[\{1\}|\{0\}]=0.1$$

$$P_C[\{0\}|\{1\}]=0.1,\qquad P_C[\{1\}|\{1\}]=0.9$$

而设计信源概率使得 $P_S[\{0\}]=P_S[\{1\}]=0.5$①，那么，在联合实验中单个事件的不同概率为

$$P[\{(0,0)\}]=P_C[\{0\}|\{0\}]P_S[\{0\}]=0.45$$

$$P[\{(0,1)\}]=P_C[\{1\}|\{0\}]P_S[\{0\}]=0.05$$

$$P[\{(1,0)\}]=P_C[\{0\}|\{1\}]P_S[\{1\}]=0.05$$

$$P[\{(1,1)\}]=P_C[\{1\}|\{1\}]P_S[\{1\}]=0.45$$

我们也可以在复合或组合空间上定义一些事件。

$X_0\triangleq$“$x=0$ 那个事件”　和　$X_1\triangleq$“$x=1$ 那个事件”

$Y_0\triangleq$“$y=0$ 那个事件”　和　$Y_1\triangleq$“$y=1$ 那个事件”

重写以上的信道条件概率为

$$P[Y_0|X_0]=0.9\quad \text{和}\quad P[Y_1|X_0]=0.1$$

$$P[Y_0|X_1]=0.1\quad \text{和}\quad P[Y_1|X_1]=0.9$$

那么信源概率就可以表示为

$$P[X_0]=0.5\quad \text{和}\quad P[X_1]=0.5$$

在组合实验中，以上的联合概率就变成了

$$P[X_0\cup Y_0]=P[Y_0|X_0]P[X_0]=0.45$$

$$P[X_0\cup Y_1]=P[Y_1|X_0]P[X_0]=0.05$$

$$P[X_1\cup Y_0]=P[Y_0|X_1]P[X_1]=0.05$$

$$P[X_1\cup Y_1]=P[Y_1|X_1]P[X_0]=0.45$$

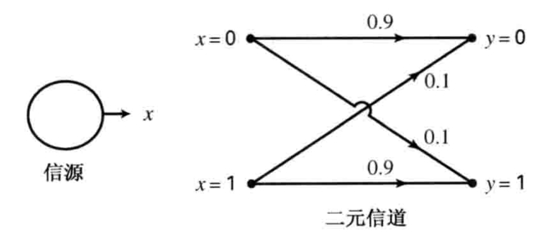


图 1.6-4　二元通信系统

条件概率的引入出现了一个重要问题，那就是条件概率是否满足公理化定理 1 ~3，换句话说，给定任意两个事件 E, F, 且 $EF=\phi$，以及任意的第三个事件 A，且 $P[A]>0$，它们都属于概率空间$(\Omega, \mathscr{F}, P)$中事件 $\mathscr{F}$ 的 σ 域，下列关系是否满足？

$$P[E|A]\geqslant 0?$$

$$P[\Omega|A]=1?$$

$$P[E\cup F|A]=P[E|A]+P[F|A],\qquad EF=\phi?$$

① 设计一个 0 和 1 出现的速率相同的编码是一个很好的练习，因为这可以使信道容量得到最大的利用。

答案是肯定的，详细的证明留给读者作为习题。它们直接根据条件概率的定义及三个科尔莫戈罗夫公理得出。

例 1.6-11 (概率树)三个事件 A，B 和 C 通常根据下列条件概率来描述：

$$P[A], P[B|A], P[B|A^c] \text{ 和}$$
$$P[C|BA], P[C|BA^c], P[C|B^cA], P[C|B^cA^c]$$

这种情况可以用图 1.6-5 所示的树图来概括。其中的分支用相关的条件概率来标注，节点值是对应的联合概率。根节点可以看成 1.0 的值，它与必然事件 Ω 相联系，如果我们要计算在树叶上事件的概率(节点的最后事件)，只需要乘以路径上的条件概率。

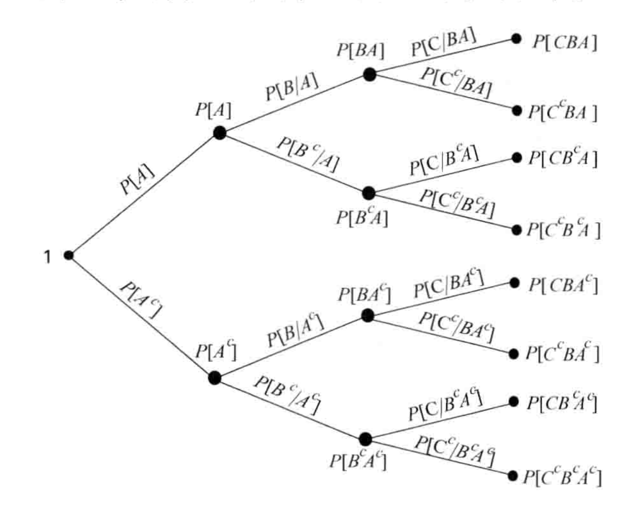


图 1.6-5 分支上标有条件概率、节点标有联合概率的概率树图

一种可能的方式是，如果事件来自于顺序执行的复合实验，即事件 B 依赖于事件 A，事件 C 依赖事件 A 和 B。更一般的树是在每个节点有两个以上的分支，如 A_1，A_2，…，A_N。条件概率可以数据结构的形式存储在机器中，联合概率的计算问题通过查询就可以得到，如联合事件 $C_kB_lA_n$ 的概率可以通过追踪存储在数据结构中的路径、然后乘以对应分支上的值即可得到。例如，第一轮事件 A_n 来表示从当地苗圃购买的植物的健康状态(好，一般，差)，然后 B_l 表示一星期后的健康状态，C_k 表示两周后的健康状态。

下一个例子用来说明联合概率和条件概率的应用，这个例子在现实生活中有广泛的应用，比如在不知道任何事实的情况下要做出重要的决定。

例 1.6-12 (选美比赛①)假定选美比赛由下列规则进行评判：(1)有 N 个竞选者，选美前评委是没有见过的。(2)竞选者以随机的顺序单独展示在评委面前，在任何时候只有一名竞选者出现在评委的面前。(3)评委必须对出现面前的竞选者是否是最美的要现场做出决定。如果评委们做出了肯定的评定，那么选美就结束了，那么风险是仍有更漂亮的竞选者没有展示出来。这样的话，评委就做出了错误的评判。换句话说，如果评委忽略了候选者，那么这个竞选者也就取消了进一步考虑的资格，即使随后所有的竞选者一点也不美。在这样的随机选择中，什么样的策略才能提高选到最美竞选者的概率？

① 感谢 Geof Williamson 和 Jerry Tiemann 对这个问题有价值的讨论。

解　为使问题更好地进行定量分析，假定每个竞选者的优点都用一个单一的“美”的数字来总结，那么，最美的竞选者就与最大的数字联系起来，最不美的就是最小的数字。我们对数字出现的分布不做任何假定。将数字放在一个包里，评委是看不到的，数字一个一个地从包里抽出。这样，就把问题看成从包里随机地抽出一个“美”数字，可以认为抽取是沿着线的顺序进行的，如图 1.6-6 所示。第一次抽取是数 1，第二次是 2，以此类推，每一次抽取，都有一个“美”数字出现，在所有 N 次抽取中哪一个最大？

假定下面采用“等等再看”(wait-and-see)的策略，我们忽略前 k 次抽取(即拒绝了前 k 个竞选者)，但记录这一组抽取的最大的“美”数字。然后继续抽取一个数字(即叫下一个竞选者出来)。在忽略前 k 次抽取后的抽取中，第一个超过前 k 次中的最大“美”数字的竞选者便是获胜者，如果没有更大的数字出现，那么评委拒绝投票，我们把这记为一个错误。

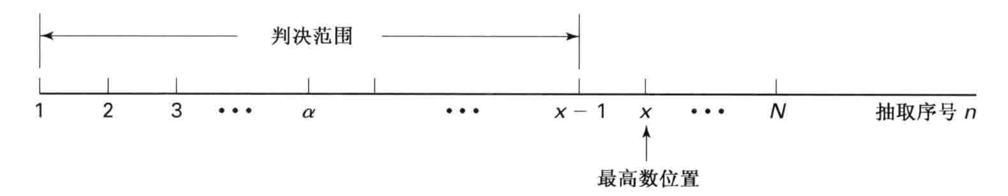


图 1.6-6　沿轴方向的数字表示抽取的顺序，并非从包里实际抽取的数字

令 $E_j(k)$ 表示前 j 次抽取的最大“美”数字出现在前 k 次抽取中这样一个事件，那么，对于 $j \leqslant k$，$E_j(k) = \Omega$(必然事件)，但对于 $j > k$，$E_j(k)$ 是 Ω 的子集。令 x 表示在包内的 N 个数字中包含有最大“美”数字这样的抽取，那么要实现一次正确的判定，两个事件必须联合发生：(1)(很明显)$\{x > k\}$；(2)(很微妙)对所有的 j，$k < j < x$，$E_j(k)$。那么，为了要得到一次正确的判定 C，子事件 $\{x = j + 1\}$ 必须与事件 $E_j(k)$(对于每一个 j，$k < j < N$)联合发生。这样，事件 $\{x > k\}$ 可以分解为一些不相交的事件的并

$$\{x > k\} = \{x = k + 1\} \cup \{x = k + 2\} \cup \cdots \cup \{x = N\}$$

因此

$$C = \{x = k + 1, E_k(k)\} \cup \{x = k + 2, E_{k+1}(k)\} \cdots \cup \{x = N, E_{N-1}(k)\}$$

正确判定的概率是

$$\begin{aligned} P[C] &= \sum_{j=k}^{N-1} P[x = j + 1, E_j(k)], \qquad \text{因为这些事件是不相交的} \\ &= \sum_{j=k}^{N-1} P[E_j(k) | x = j + 1] P[x = j + 1] \\ &= \frac{1}{N} \sum_{j=k}^{N-1} \frac{k}{j} \end{aligned}$$

其中利用了这样一个事实 $P[x = j + 1] = 1/N$，这是由于从 N 次抽取中得到最大“美”数字是等可能的。另外，$P[E_j(k) | x = j + 1] = \dfrac{k}{j}$，是因为从前 j 次抽取中任何一次抽到最大“美”数字是等可能的，所以，在 j 次抽取中的前 k 次抽得的概率为 k/j。

由欧拉(Euler)求和公式①,对于大的 N

$$
\begin{aligned}
P[C] &= \frac{1}{N}\sum_{j=k}^{N-1}\frac{k}{j} \\
&= \frac{k}{N}\left(\frac{1}{k}+\frac{1}{k+1}+\frac{1}{k+2}+\cdots+\frac{1}{N-1}\right) \\
&\approx \frac{k}{N}\int_k^N \frac{\mathrm{d}x}{x} \qquad \text{,当} k \text{ 足够大时} \\
&= \frac{k}{N}\ln\frac{N}{k}
\end{aligned}
$$

忽略整数约束,k 的近似的最佳选择(比如说 k_0)可通过微分求得,令

$$\frac{\mathrm{d}P[C]}{\mathrm{d}k}=0$$

我们求得

$$k_0 \approx \frac{N}{e}$$

借助整数的约束,我们把 k_0 四舍五入,得

$$k_0 \approx \left\lfloor \frac{N}{e}+\frac{1}{2}\right\rfloor$$

其中$\lfloor \cdot \rfloor$表示取最小的整数。正确判定的最大概率 $P[C]$ 变成

$$
\begin{aligned}
P[C] &\approx \frac{\lfloor \frac{N}{e}+\frac{1}{2}\rfloor}{N}\ln\frac{N}{\lfloor \frac{N}{e}+\frac{1}{2}\rfloor} \\
&\approx \frac{1}{e}\ln e \doteq 0.367
\end{aligned}
$$

这样,我们在开始对竞选者做出判定之前,应该近似让前三个竞选者(更准确地说是 36.7%)过去。假定 N 是这一结论成立的一个合理的大数字,有趣的事实是这个结论与 N 无关,而通过随机选择选到最美候选人的概率则随 $1/N$ 减少。

还有其他一些情况需要使正确判决概率最大的策略。

1. 你是一个要租房子的人,选择了 30 个出租公寓进行考察,你看到了一个你喜欢的公寓,但没有签约,因为你认为下一个要看的公寓或许更是你所希望的,然而,随后你看到的公寓没有一个比得上前一个。在此期间,你拒绝的那套公寓被别人租了,由于你的犹豫,不得不租一套差的公寓。
2. 你正在寻找一位伴侣以度过你的余生,为此你与单身约会机构接触,以每周一次约会的速度去与 50 位可能的伴侣见面,在第 9 次约会时,你决定你找到了生命中的伴侣,并且与其结婚。然后,你忘记通知约会机构停止介绍其他人约会。下一周,又介绍你去约会,而这个新介绍的可能伴侣各方面的条件都优于你已选择的,你为你的冲动而猛踢自己。

① 例如,参见 G. F. Carrier 等的 *Function of a Complex Variable*(New York: McGraw-Hill, 1966), p. 246 或访问维基网页: Euler-Maclaurin formula (http://en.wikipedia.org/wiki/Euler%E2%80%93Maclaurin_formula)。

3. 你在为政府高级职位的候选人进行面试。为了减少偏心的可能性，你受到下列规则的约束：你对候选人的面试是按顺序来进行的，并且根据工作的描述，把工作给第一个合格的候选人。如果你拒绝候选人，这意味着他/她是不合格的，那你必须在你的报告中进行陈述。然而，你足够聪明，知道即使在合格的候选人中还有更为出色的人选，而其他人仅仅是合格而已。你要为工作雇用最佳人选。你的策略是什么？

全概率定理　在工程和科学的许多问题中，我们要计算根据加权条件概率之和来计算事件 B 的无条件概率 $P[B]$，这样的计算很容易通过下面的定理来实现。

定理 1.6-1　令 A_1, A_2, $\cdots$, A_n 是 n 个互斥的事件，且 $\cup_{i=1}^{n}A_i=\Omega$（A_i 是穷举的），令 B 是任何一个定义在 A_i 的概率空间上的事件，那么，对于所有的 i，$P[A_i]\neq 0$，有

$$P[B]=P[B|A_1]P[A_1]+\cdots+P[B|A_n]P[A_n] \tag{1.6-7}$$

有时 $P[B]$ 也称为 B 的全概率，因为右边的表达式是 B 的条件概率的加权平均。

证明　对于所有的 $i\neq j$，我们有 $A_iA_j=\phi$ 和 $\cup_{i=1}^{n}A_i=\Omega$，另外，$B\Omega=B=B\cup_{i=1}^{n}A_i=\cup_{i=1}^{n}BA_i$，但是，由交集运算的定义，$BA_i\subset A_i$，因此，所有的 $i\neq j$，有 $(BA_i)(BA_j)=\phi$，这样，由公理 3（推广到 n 个事件）

$$\begin{aligned}P[B]&=P\left[\bigcup_{i=1}^{n}BA_i\right]=P[BA_1]+P[BA_2]+\cdots+P[BA_n]\\&=P[B|A_1]P[A_1]+\cdots+P[B|A_n]P[A_n]\end{aligned} \tag{1.6-8}$$

最后一行是由式(1.6-2)得出的。

例 1.6-13　（二元信道的更多讨论）对于图 1.6-4 所示的二元通信系统，计算无条件输出概率 $P[Y_0]$ 和 $P[Y_1]$。

解　继续二元通信的例 1.6-10 的符号表示，利用式(1.6-8)，得

$$\begin{aligned}P[Y_0]&=P[Y_0|X_0]P[X_0]+P[Y_0|X_1]P[X_1]\\&=P_C[0|0]P_S[0]+P_C[0|1]P_S[1]^{\dagger}\\&=(0.9)(0.5)+(0.1)(0.5)\\&=0.5\end{aligned}$$

我们可以用类似的方式计算 $P[Y_1]$，或者注意到 $Y_0\cup Y_1=\Omega$ 和 $Y_0\cap Y_1=\phi$；即它们是不相交的，因此，$P[Y_0]+P[Y_1]=1$，这也意味着 $P[Y_1]=1-P[Y_0]=0.5$。

1.7　贝叶斯定理及应用

前面的结果使我们能够用简单的公式写出，这个公式称为贝叶斯定理①。尽管很简单，这个公式在生物统计学、流行病学和通信理论中有广泛的应用。

贝叶斯定理　令 A_i（$i=1, 2, \cdots, n$）是定义在概率空间 $\mathscr{P}$ 上的一个互不相交且穷举事件的集合，那么，对于所有的 $i\neq j$，$\cup_{i=1}^{n}A_i=\Omega$，$A_iA_j=\phi$。有一个定义在 $\mathscr{P}$ 上的事件 B，$P[B]>0$ 和对所有的 i，$P[A_i]\neq 0$，有

① 以英国数学家、哲学家 Thomas Bayes 命名。

$$P[A_j|B] = \frac{P[B|A_j]P[A_j]}{\sum_{i=1}^{n} P[B|A_i]P[A_i]} \tag{1.7-1}$$

证明 分母是由式(1.6-8)给出的$P[B]$，分子是$P[A_j B]$，因此，贝叶斯定理只是条件概率的应用。

注释 在实际中，给式(1.7-1)的各项不同的名字，$P[A_j|B]$称为B给定的条件下A_j的后验概率，$P[A_i]$是A_i的先验概率。一般说来，先验概率是根据过去的测量得到的估计或依据经验的预先假定，而后验概率是根据观测来进行度量的或计算的。

例 1.7-1 (逆二元信道)在通信系统中，0或1分别以概率$P_S[0]=p_0$，$P_S[1]=1-p_0 \triangleq p_1$被发射，由于信道中有噪声，0被接收为1的概率为$\beta$，称为交叉概率，1被接收为0的概率也为$\beta$。在信道的输出端观测到1，源的输出或信道的输入(即发射)为1的概率是多少？

解 信道的结构如图1.7-1所示，可写成

$$P[X_1|Y_1] = \frac{P[X_1Y_1]}{P[Y_1]} \tag{1.7-2}$$

$$= \frac{P_C[1|1]P_S[1]}{P_C[1|1]P_S[1] + P_C[1|0]P_S[0]} \tag{1.7-3}$$

$$= \frac{p_1(1-\beta)}{p_1(1-\beta) + p_0\beta} \tag{1.7-4}$$

如果$p_0=p_1=1/2$，逆概率或后验概率$P[X_1|Y_1]$依赖于β，如图1.7-2所示，如果$\beta=0$，我们称信道是无噪声的，但注意到当$\beta=1$时，信道一样有用，在这种情况下，输出刚好是反过来的。

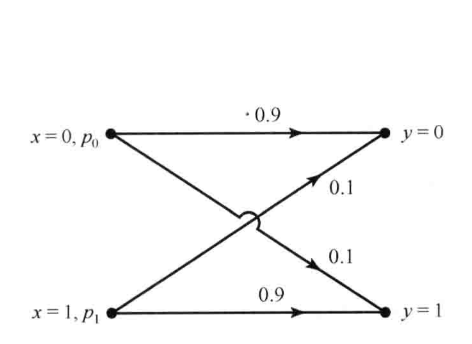


图 1.7-1 有噪声时二元通信信道的表示

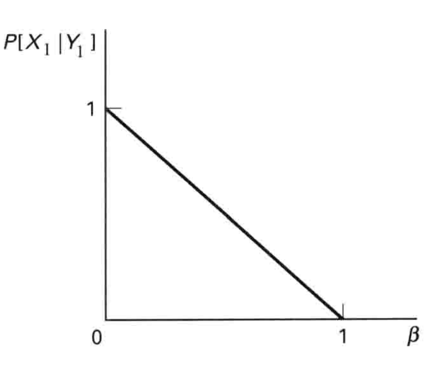


图 1.7-2 后验概率与β的关系

例 1.7-2 [阿尔茨海默(Alzheimer)病的淀粉样蛋白测试]2010年8月10日，在网络电视新闻中有一个关于阿尔茨海默病的一种有前途的新的测试，它是基于在脊柱(脑)流体中淀粉状蛋白的出现，它可以通过脊椎抽液检测到。据报道，在老年痴呆症患者中(65岁以上)，有90%的病人有淀粉状蛋白，而在无老年痴呆症的人组中(65岁以上)，只有36%的有淀粉状蛋白。从过去几年的调查得出的结论是：在65岁以上的人组中，老年痴呆一般的发病率是10%。从这个数据我们要求证：这真的是一个好测试？

首先构造这个测试的概率空间，令$\Omega=\{00, 01, 10, 11\}$，有四个实验结果：

00 =“没有淀粉状蛋白”和“没有老年痴呆症”，

01 =“没有淀粉状蛋白”和“有老年痴呆症”，

10 =“有淀粉状蛋白”和“没有老年痴呆症”，

11 =“有淀粉状蛋白”和“有老年痴呆症”。

在这个样本空间中，我们定义了两个事件：$A \triangleq \{10, 11\}$ =“有淀粉状蛋白”和 $B \triangleq \{01, 11\}$ =“有老年痴呆症”，从以上的数据，我们有

$$P[A|B] = 0.9 \text{ 和 } P[A|B^c] = 0.36$$

另外，根据总的样本数(65 个以上)，我们有

$$P[B] = 0.1 \qquad \text{和} \qquad P[B^c] = 1 - P[B] = 0.9$$

现在要确定测试是不是好的，我们必须看一看做了测试以后的情况，这可以用测试后的条件概率来建模，即如果测试淀粉状蛋白是阳性的就用 $P[\cdot|A]$，如果测试是阴性的就用 $P[\cdot|A^c]$。所以，可以用全概率公式来求得 $P[B|A]$ 的答案，于是

$$\begin{aligned} P[B|A] &= \frac{P[A|B]P[B]}{P[A|B]P[B] + P[A|B^c]P[B^c]} \\ &= \frac{0.9 \times 0.1}{0.9 \times 0.1 + 0.36 \times 0.9} \\ &= 0.217 \end{aligned}$$

因此，在测试为阳性的组中，实际上只有22% 有老年痴呆症。现在看来这个测试并不是那么有前途，为什么呢？问题就在于，在涉及到条件概率 $P[\cdot|B]$ 时，我们永远也没有由事件 B 刻画的“知识状态”，在给定测试前，我们的知识状态是由无条件概率知识 $P[\cdot]$ 来刻画的。测试以后，知识状态是由事件 A 或 A^c 是否已经发生来刻画的。即我们的条件概率是 $P[\cdot|A]$ 或者 $P[\cdot|A^c]$。你看到的是，我们进入了检验人群的“给定 A”或者“给定 A^c”的知识状态中，所以，永远也不知道 $P[\cdot|B]$ 或 $P[\cdot|B^c]$ 这种相关的概率测度的知识状态。所以，给定信息 $P[A|B] = 0.9$ 和 $P[A|B^c] = 0.36$ 对直接确定测试是有用还是无用是没有什么帮助的。这是用 $P[A|B]$ 代替 $P[B|A]$ 进行逻辑推理的一个谬论，但在这个特殊的例子里还有非常实际的问题要继续讨论。

当我们计算 $P[B^c|A] = 1.0 - 0.217 = 0.783$ 时，这意味着淀粉状蛋白测试为阳性的人的78% 并没有老年痴呆，所以，由于假阳性率太高，这个测试是无用的。另外，从前一个例子，在总的样本数(65 个以上)中患有老年痴呆症的人数稀少也是一个问题，任何测试要成为有用的测试都必须克服这个问题。

1.8　组合①

在继续概率基础学习之前，对于等概事件的计数，引入一些重要的计数公式。这里给出的结果有些将在 1.9 节中立即得到应用，其他在后面也会有应用。

容量为 n 的总体意味着是一个 n 个元素的集合，元素在集合中是不考虑顺序的。来自于容量为 n 的总体中的一个容量为 r 的子总体是一个 r 个元素的子集，这 r 个元素是来自于原始总体的。同样，两个子总体认为是不同的，指的是一个子总体至少有一个元素不同于另一个子总体中的元素。

考虑一个 n 个元素 $a_1, a_2, \cdots, a_n$ 的总体，r 个符号的任何有序的排列 $a_{k_1}, a_{k_2}, \cdots, a_{k_r}$ 称

① 这一部分材料采用了 Willian Feller[1-8]的相关内容。

为容量为 r 的有序样本。现在考虑通用罐，罐里装有 n 个标有数字的球，球一个接一个地被拿出，这能形成多少个容量为 r 的不同的有序样本？有两种情况：

(i) 重复抽样　每个球取出后，记录球的数字，然后放回罐里。这样，第一次取样有 n 种选择，第二次取样还是 n 种选择，以此类推。这样就得到如下结果：对于一个 n 个元素的总体，能够形成 n^r 个容量为 r 的不同的有序样本。

(ii) 无重复抽样　每个球被抽出后不再放回原处，随后的抽样就不可用。这样，第一次抽样，有 n 个球可用，第二次抽样就只有 $n-1$ 个球可用，以此类推。于是，就得到这样的结果：对于 n 个元素的总体，通过无重复抽样，形成了

$$(n)_r \triangleq n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} \tag{1.8-1}$$

个容量为 r 的不同的有序样本①。

在大小为 n 的总体中大小为 r 的子总体的个数在概率论中常常出现的一个基本问题是：从一个容量为 n 的总体中能够形成多少个组或形成多少子总体？例如，考虑标有数字 1 ~6 的 6 个球，能够形成多少个容量为 2 的组？下面的表格表明可以形成 15 个容量为 2 的组：

12	23	34	45	56
13	24	35	46	
14	25	36		
15	26			
16				

注意，这不同于无重复抽样能够形成的有序样本组的数目，这些样本组是(6 · 5 =30)

12	21	31	41	51	61
13	23	32	42	52	62
14	24	34	43	53	63
15	25	35	45	54	64
16	26	36	46	56	65

另外，这也不同于重复抽样能够形成的样本组数目($6^2=36$)

11	21	31	41	51	61
12	22	32	42	52	62
13	23	33	43	53	63
14	24	34	44	54	64
15	25	35	45	55	65
16	26	36	46	56	66

在容量为 n 的总体中，容量为 r 的子总体数 C_r^n 的通用公式能够计算如下：考虑一个有 n 个可识别的球，我们已经知道，能够形成的容量为 r 的有序样本数目是 $(n)_r$，现在考虑一个容量为 r 的特殊子总体，这个子总体有 $r!$ 个排列，也就有 $r!$ 个不同的有序样本。所以，对于 C_r^n 个子总体，肯定有 $C_r^n \cdot r!$ 容量为 r 的不同有序样本，因此

$$C_r^n \cdot r! = (n)_r$$

或

① 不同的样本常常包含有相同的子总体，但具有不同的顺序，由于这个原因，我们说无重复抽样能形成的有序样本组的数目是 $(n)_r$ 个。

$$C_r^n=\frac{(n)_r}{r!}=\frac{n!}{(n-r)!r!}\triangleq\binom{n}{r} \tag{1.8-2}$$

式(1.8-2)是一个很重要的结果，我们将在下一节应用。符号

$$C_r^n\triangleq\binom{n}{r}$$

称为二次项系数。很显然

$$\binom{n}{r}=\frac{n!}{r!(n-r)!}=\frac{n!}{(n-r)!r!}\triangleq\binom{n}{n-r}=C_{n-r}^n \tag{1.8-3}$$

从 1.4 节已经知道，容量为 n 的集合的子集的数目是 2^n，容量为 r 的子集的数目是$\binom{n}{r}$，因此，我们得到

$$\sum_{r=0}^{n}\binom{n}{r}=2^n$$

下面的定理给出一个结论，这个结论可以看成二项式系数 C_r^n 的扩展。

定理 1.8-1　令 r_1，…，r_l 是 l 个非负整数的集合，$r_1+r_2+\cdots+r_l=n$，那么 n 个元素的总体划分为 l 个子总体的方法有

$$\frac{n!}{r_1!r_2!\cdots r_l!} \tag{1.8-4}$$

其中这些子总体的第一个含有 r_1 个元素，第二个含有 r_2 个元素，以此类推。这个系数称为多项式系数。注意，子总体是有顺序的，即($r_1=7$，$r_2=10$)和($r_1=10$，$r_2=7$)表示不同的划分。然而，每个组里的顺序并不重要。例如，假定有五个可分辨的球(1，2，3，4，5)，我们要问能够划分出多少个子总体？其中第一组有三个球，第二组有两个球。即 $n=5$，$r_1=3$，$r_2=2$，且 $r_1+r_2=5$，答案是 $5!/3!2!=10$，这些划分是

第一组:	1,2,3	2,3,4	3,4,5	4,5,1	5,1,2	2,4,5	2,3,5	1,3,5	1,3,4	1,2,4
第二组:	4,5	5,1	1,2	2,3	3,4	1,3	1,4	2,4	2,5	3,5

注意，这里顺序是重要的，如果令 $r_1=2$，$r_2=3$，那我们就会得到不同的划分，例如

第一组:	4,5	5,1	1,2	2,3	3,4	1,3	1,4	2,4	2,5	3,5
第二组:	1,2,3	2,3,4	3,4,5	4,5,1	5,1,2	2,4,5	2,3,5	1,3,5	1,3,4	1,2,4

然而划分(4，5)，(1，2，3)与(5，4)，(2，1，3)是相同的。

证明：注意到我们能够重写式(1.8-4)为

$$\frac{n!}{r_1!}\cdot\frac{1}{r_2!}\cdot\cdots\cdot\frac{1}{r_l!}$$

$$=\frac{n!}{r_1!(n-r_1)!}\cdot\frac{(n-r_1)!}{r_2!(n-r_1-r_2)!}\frac{(n-r_1-r_2)!}{r_3!(n-r_1-r_2-r_3)!}\cdots\frac{\left(n-\sum_{j=1}^{l-1}r_j\right)!}{r_l!\left(n-\sum_{j=1}^{l}r_j\right)!}$$

回想一下，$0!=1$，可看到最后一项是 1，那么，多项式公式写为

$$\frac{n!}{r_1!r_2!\cdots r_l!}=\binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\cdots\binom{n-r_1-r_2-\cdots-r_{l-2}}{r_{l-1}} \tag{1.8-5}$$

为了影响在第一子总体的 r_1 个元素的现实，在第二个子总体中的 r_2 个元素的现实，等等，我们应该从给定的 n 个元素中选择 r_1 个元素，从剩下的 $n-r_1$ 中选择 r_2 个元素，从剩下的 $n-r_1-r_2$ 中选择 r_3 个元素，等等。但是，在 n 中选择 r_1 个元素有 $\binom{n}{r_1}$ 种选法，从 $n-r_1$ 中选 r_2 个元素有 $\binom{n-r_1}{r_2}$ 种选法，以此类推。因此，这样选择的话，总的选法就是式(1.8-5)右边各项之积，证明完成。

例 1.8-1 (掷 12 个骰子)[1-8, p.36] 假定我们掷 12 个骰子，每个骰子有 6 个实验结果，总共有 $n_T=6^{12}$ 种实验结果。现在考虑每面出现两次事件 E，当然，这个事件的发生有很多种方式，两个实验结果发生如下：

骰子I.D.号:	1	2	3	4	5	6	7	8	9	10	11	12
实验结果1	3	1	3	6	1	2	5	4	4	6	2	5
实验结果2	6	1	3	2	6	3	4	5	5	1	4	2

这个事件发生的总次数就是 12 个骰子($n=12$)被安排成 6 组($k=6$)、每组两个($r_1=r_2=\cdots=r_6=2$)的总数。假定所有实验结果是等可能的，我们得

$$P[E]=\frac{n_E}{n_T}=\frac{E\text{ 发生的次数}}{\text{总的实验结果数}}=\frac{12!}{(2!)^6 6^{12}}=0.003\ 438$$

二项式和多项式系数出现在下一节将要讨论的二项式分布和多项式分布。多项式系数在理论物理中称为入住问题中也是很重要的。

入住问题*

入住问题本质上可以看成把 r 个球随机放入 n 个单元的问题，第一个球有 n 种选择，第二个球也有 n 种选择，以此类推，所以，在 n 个单元中 r 个球有 n^r 种可能的分配方法，每一种方法的概率是 n^{-r}，如果球是可辨别的，那么，每一种分配也是可辨别的，如果球不是可辨别的，那么，可辨别的分配就少于 n^r 种。例如，用三个可辨别的球($r=3$)，分别标有“1”，“2”，“3”，有两个单元($n=2$)，那么，就可以得到 2^3 种可辨别的分配：

单元1	1	2	3	1,2	1,3	2,3	1,2,3	—
单元2	2,3	1,3	1,2	3	2	1	—	1,2,3

当球是不可辨别的，每个球用一个“ * ”表示，我们可得到四个不同的分配：

单元1	***	**	*	—
单元2	—	*	**	***

从 r 个球和 n 个单元能形成多少种可辨识的分配？威廉 · 费勒[1-8, p.38] (William Feller) 提供一种简便的方法，这种方法使用了很聪明的技巧，技巧是在 $n+1$ 个竖线之间的空间表示 n 个单元，并且用星号“ * ”表示球，这样

| | | |

表示三个空单元，而

|**| | |*|**| |*****|

表示在第一个单元中有两个球，第二、三个单元中有零个球，第四个单元中有一个球，第五个单元中有两个球，等等。实际上，用 $r_i\geqslant 0$ 表示在第 i 个单元中球的个数，r 表示总的球数，那么

$$r_1+r_2\cdots r_n=r$$

n 元组(r_1, r_2, …, r_n)称为入住，r_i 是入住数，在单元中球的两种分配，如果它们对应的入住是相同的，则称为不可辨识。入住

|**|　|　|*|**|　|*****|

是(2, 0, 0, 1, 2, 0, 5)。注意，n 个单元要求 $n+1$ 根竖线，因为第一符号和最后一个符号必须是竖线。$n-1$ 根竖线和 r 个星号“*”可以任何顺序出现。因此我们要问，在容量为 $n-1+r$ 个总体中，容量为 r 的子总体的数量是多少呢？由式(1.8-2)得出的结论是

$$\binom{n+r-1}{r}=\binom{n+r-1}{n-1} \tag{1.8-6}$$

例 1.8-2　(可辨识分配)证明没有一个单元是空的可辨识分配是 $\binom{r-1}{n-1}$，即我们要求没有一根竖线是相邻的。因此，r 个“*”必须入住竖线之间的空间，剩下的 $r-n$ 个“*”可以放入任何位置。这样，$n-1$ 根竖线和 $r-n$ 个“*”可以任何顺序出现，那么，不同的分配数等于选择 $r-n$ 个“*”放置到($n-1$)根竖线之间、再加上($r-n$)个星号放置到任意的竖线之间的总的选择数，即 $n-1+r-n=r-1$ 中选 $r-n$ 的总的选择数。于是，由式(1.8-2)，得

$$\binom{r-1}{r-n}=\binom{r-1}{n-1}$$

例 1.8-3　(同一天生日)小组的人惊讶地发现他们的生日与组里的其他人经常一致，在宣布这一神秘事情之前，我们先把它当成一个入住问题进行分析。即我们要计算需要多大的一组才有高概率出现生日“碰撞”，即在组内至少有两个人在同一天生日。

解　令一年的天数 $n(n=365)$ 用 n 个单元表示，一组里 r 个人用 r 个球表示，那么，当球放入一个单元时，也就是球所代表的人在那天生日，当一个单元有两个以上的球时就表示生日“碰撞”发生，现在考虑球的排列。第一个球可以放置在 n 个单元中的任意一个，但第二个球只有 $n-1$ 种选择才能避免“碰撞”，第三个球就只有 $n-2$ 种选择，按这种方式继续下去，我们发现，避免“碰撞”的排列数是 $n(n-1)\cdots(n-r+1)$。在 n 个单元中 r 个球的总的排列数是 n^r。因此，用 $P_0(r, n)$ 表示零次生日“碰撞”的概率与 r 和 n 的函数，可求得为 $P_0(r, n)=\dfrac{n(n-1)\cdots(n-r+1)}{n^r}=\prod_{i=1}^{r-1}\left(1-\dfrac{i}{n}\right)$，那么 $1-P_0(r, n)$ 就是至少有一次“碰撞”的概率。

需要多大的 r 才能使 $1-P_0(r, n)\geqslant 0.9$ 呢？或等价于 $P_0(r, n)\leqslant 0.1$？剩下的问题就是 $\prod_{i=1}^{r-1}\left(1-\dfrac{i}{n}\right)\leqslant 0.1$ 的求解。我们使用基本微积分中的一个结论：对任意实数 x，$1-x\leqslant \mathrm{e}^{-x}$，在 x 接近零时，这是一个很好的近似。如果用 $\prod_{i=1}^{r-1}\mathrm{e}^{-\frac{i}{n}}\leqslant 0.1$ 替代 $\prod_{i=1}^{r-1}\left(1-\dfrac{i}{n}\right)\leqslant 0.1$，那就得到一个边界和 r 的估计。由于 $\prod_{i=1}^{r-1}\mathrm{e}^{-\frac{i}{n}}=\exp\left\{-\dfrac{1}{n}\sum_{i=1}^{r-1}i\right\}$，以及 $\sum_{i=1}^{r-1}i=r(r-1)/2$，那么得出的表达式 $\mathrm{e}^{-\frac{1}{2n}r(r-1)}\leqslant 0.1$ 将给出 r 的估计值。假定 $r^2\gg r$，$n=365$，我们得到 $r\approx 40$。所以，在 40 人一组中至少有两个人同一天生日的概率为 90%。

例 1.8-4　(“十三点”游戏)在十七世纪的威尼斯，狂欢节期间，带着喜剧中角色的面具的赌徒在称为化妆舞会的娱乐大厅玩着一种纸牌游戏“十三点”。

在"十三点"游戏中，一名游戏者扮演银行家，其他游戏者将放置赌注，银行家在台面上洗牌，牌面向下，然后按顺序从 1～13 叫出 A 到 K 的牌的名字，每次银行家翻出一张牌，如果翻出的牌与他叫的数字匹配，他就赢了，并且收走所有赌注，如果翻出的牌与银行家叫出的数字不匹配，游戏继续，直到庄家叫到 13，如果翻过来的牌不是 K，银行家就输掉本次游戏，他就必须给每位下注者赔付对方下的注，这时，扮演银行家的游戏者必须让出他的位置，把银行家的角色让给右边的游戏者。

银行家赢得比赛的概率是多少呢?

解　为了简化分析，假定一次只翻一张牌，如果不匹配就放回原处，然后重新洗牌，再翻下一张牌。令 A_n 表示银行家第一次匹配这个事件，即在第 n 次翻牌时赢了，用 W_n 表示在 n 次翻牌中有一次赢这个事件。由于在 52 张牌中每个数字有四张牌，匹配的概率是 1/13，为了要在第 n 次翻牌时第一次赢得游戏，就应该在前 $n-1$ 次不匹配然后跟着一次匹配，这一事件的概率为

$$P[A_n] = \left(1-\frac{1}{13}\right)^{n-1}\frac{1}{13}$$

由于对于 $i \neq j$，$A_iA_j = \phi$，所以，赢得"十三点"游戏的概率是

$$P[W_{13}] = \sum_{i=1}^{13} P[A_i] = \frac{1}{13}\frac{1-\left(\frac{12}{13}\right)^{13}}{\left(1-\frac{12}{13}\right)} = 0.647$$

而银行家输掉游戏这个事件 W_{13}^c 的概率是 $P[W_{13}^c]=0.353$。实际上，利用每次匹配的概率 1/13，通过考察银行家在 13 次翻牌中没一次匹配就输掉游戏这个事件，那么输掉游戏的概率也是更容易得到的

$$P[W_{13}^c] = \binom{13}{0}\left(\frac{1}{13}\right)^0\left(\frac{12}{13}\right)^{13} = 0.353$$

注意，在第二个等式中，使用了几何级数求和公式：$\sum_{n=0}^{N-1} x^n = \frac{1-x^N}{1-x}$ (参见附录 A)。

要考虑的几点：为什么当 $n \to \infty$ 时，$P[A_n] \to 0$? 为什么对所有的 n，$P[W_n] \geq P[A_n]$? 如果没有假定翻牌后放回原处，又如何重新建模呢? 如果银行家从 13(K) 到 1(A) 反方向叫牌，这个问题又如何变化呢?

在统计力学中，由三个位置和三个动量坐标表示的六维空间称为相空间，根据不确定性原理，位置的不确定性乘以动量的不确定性不能小于普朗克(Planck)常数 h，相空间被量化为一个 $v=h^3$ 的小的体元，在包含原子或分子大小的粒子的系统中，这些粒子在这些体元之间的分布构成了系统的状态。在麦克斯韦-玻尔兹曼统计学中，在 n 个单元中的 r 个粒子是等可能的。可以证明(参见由 A. Beiser 所著，McGraw-Hill 出版的 *Concepts of Modern Physics*)得出著名的玻尔兹曼定律：

$$n(\varepsilon) = \frac{2\pi N}{(\pi k)^{3/2}}\sqrt{\varepsilon}\ \mathrm{e}^{-\varepsilon/kT} \tag{1.8-7}$$

其中，$n(\varepsilon)\mathrm{d}\varepsilon$ 是能量在 ε 到 $\varepsilon+\mathrm{d}\varepsilon$ 之间的粒子数，N 是总的粒子数，T 是热力学温度，k 是玻尔兹曼常数。在某种意义下，对于能够辨识的同一类粒子，麦克斯韦-玻尔兹曼定律是成立的。有争论的是气体的分子是这类粒子，式(1.8-7)对 N 求积分并不困难。

与麦克斯韦－玻尔兹曼统计学不同的是，在所有 n^r 种排列等可能的时候，玻色-爱因斯坦

(Bose-Einstein)统计学只考虑了不可辨识的相同粒子的可辨识的排列。对于 n 个单元和 r 个粒子，这样的排列数由式(1.8-6)给定。

$$\binom{n+r-1}{r}$$

每一种排列被赋予一个概率

$$\binom{n+r-1}{r}^{-1}$$

有人认为玻色-爱因斯坦统计学对那些不服从不相容原理的零自旋或整数自旋的光子、原子核和粒子是有效的。不相容原理是由沃尔夫冈・泡利(Wolfgang Pauli)在 1925 年发现的，对于某一类粒子(如电子)，不会有两个粒子同时存在于相同的量子态(即没有两个或以上的球处于同一单元)。

为了处理服从不相容原理的粒子，分析第三种概率赋值方法，这种赋值称为费米-狄拉克(Fermi-Dirac)统计，假定

(1) 不相容原理(在同一单元中没有两个以上的球)。

(2) 所有满足(1)的可辨识的排列是等可能的。

注意，对于费米-狄拉克统计，$r\leqslant n$，在不相容原理的假设下，可辨识排列数是在 n 个元素的总体中容量为 $r(r\leqslant n)$ 的子总体的个数，即$\binom{n}{r}$，由于每一种排列是等可能的，任何一种状态的概率是$\binom{n}{r}^{-1}$。

以上讨论应该使读者相信，在基础科学中概率的重要性以及局限性。没有纯粹的基于概率公理的推理能够确定哪些粒子服从哪些概率定律。

扩展与应用

定理 1.5-1 关于事件并的概率能够用来解决工程中问题，首先，我们注意到在和 S_i 中学的概率项数是$\binom{n}{i}$，为什么呢？因为在 S_i 中有 n 个索引，所有项都有 i 个索引。例如，对于 $n=5$ 和 $i=2$，S_2 将由项 P_{ij}的和组成，其中，索引 ij 是 12；13；14；15；23；24；25；34；35；45。在 S_i 中的每个索引集永远也不会重复，即它们是不同的。这样，索引数(也即是 S_i 中的项数)是容量为 n 的总体中容量为 i 的子总体的个数，根据式(1.8-2)，这个数是$\binom{n}{i}$。注意，S_n 只有一个单项。

例 1.8-5　给定 r 个球和 n 个单元，球是不可辨识的，并随机地分布在 n 个单元中，假定每个排列是等可能的，计算所有单元都有球占住的概率。注意，球可以表示数据包，而单元表示缓冲器，或者球表示空投食品配给，而单元表示处于饥荒国家的灾民。

解　令 E_i 表示单元 $i(i=1, \cdots, n)$是空的这个事件，那么，r 个球放置在剩余的 $n-1$ 单元，因为 r 个球的每一个有 $n-1$ 个单元可选择，因此把 r 个球放置到 $n-1$ 个单元有 $A(r, n-1)\triangleq(n-1)^r$ 种排列方法。很明显，由于球是不可辨识的，并非所有的排列都是可辨识的，实际上，只有$\binom{n+r-1}{n-1}$可辨识的分配，通常，这些分配不是等可能的。将 r 个球分配到 n 个

单元的总数是 n^r，因此

$$P[E_i]=(n-1)^r/n^r=\left(1-\frac{1}{n}\right)^r \triangleq P_i$$

下一步假定单元 i 和 j 是空的，那么，$A(r, n-2)=(n-2)^r$

$$P[E_iE_j]\triangleq P_{ij}=\left(1-\frac{2}{n}\right)^r$$

用类似的方法，很容易证明，$P[E_iE_jE_k]=\left(1-\frac{3}{n}\right)^r\triangleq P_{ijk}$，以此类推。注意，$P_i$，$P_{ij}$，$P_{ijk}$等右边表达式，并没有包含下标 i，ij，ijk，等等。于是，每一个 S_i 包含$\binom{n}{i}$个相同项，它们的和为

$$S_i=\binom{n}{i}\left(1-\frac{i}{n}\right)^r$$

令 E 表示至少有一个单元是空的，那么，由定理 1.5-1 可得

$$P[E]=P\left[\bigcup_{i=1}^{n}E_i\right]=S_1-S_2+\cdots+S_n$$

从上面两行的 S_i 代入，得

$$P[E]=\sum_{i=1}^{n}\binom{n}{i}(-1)^{i+1}\left(1-\frac{i}{n}\right)^r \tag{1.8-8}$$

所有单元都被占用这个事件是 E^c，则 $P[E^c]=1-P[E]$ 可以写成

$$P[E^c]=\sum_{i=0}^{n}\binom{n}{i}(-1)^{i}\left(1-\frac{i}{n}\right)^r \tag{1.8-9}$$

例 1.8-6　(m 个单元空)在 r 个球已经分配后，用式(1.8-9)精确计算 n 个单元中有 m 个单元是空的概率，我们用三参数函数 $P_m(r, n)$来表示这个概率。

解　写 $P[E^c]=P_0(r, n)$，现在假定 m 个单元是空的，$n-m$ 个单元被占用，然后固定 m 个空的单元，例如，单元$\underbrace{2, 4, 5, 7, \cdots, l}_{m项}$。令 $B(r, n-m)$是在剩下的 $n-m$ 个单元分配 r 个球以至于剩下的单元没有空的总的分配数，那么，$P_0(r, n-m)=B(r, n-m)/A(r, n-m)$，由于 $A(r, n-m)=(n-m)^r$，我们得到 $B(r, n-m)=(n-m)^rP_0(r, n-m)$，在 n 个单元中，有$\binom{n}{m}$种方法让 m 个单元空着。因此，n 个单元中在剩余 m 个单元空着的情况下，r 个球排列的总数是$\binom{n}{m}(n-m)^rP_0(r, n-m)$。最后，在 n 个单元中分配 r 个球的方法总数是 n^r，因此

$$P_m(r,n)=\binom{n}{m}(n-m)^rP_0(r,n-m)/n^r$$

化简后得

$$P_m(r,n)=\binom{n}{m}\sum_{i=0}^{n-m}\binom{n-m}{i}(-1)^i\left(1-\frac{i+m}{n}\right)^r \tag{1.8-10}$$

1.9　伯努利试验：二项式和多项式概率分布

考虑一个简单的随机实验，它是由只有两个实验结果的单一试验组成：成功$\{\zeta_1 = s\}$，概率为p，$0<p<1$，或失败$\{\zeta_2 = f\}$，概率为$q = 1-p$。因此，$P[s]=p$，$P[f]=q$，样本空间为$\Omega=\{s, f\}$，事件$\mathscr{F}$的σ域是ϕ，Ω，$\{s\}$，$\{f\}$，这个试验称为伯努利试验。

假定做两次试验，新的样本空间为Ω_2，写成$\Omega_2=\Omega\times\Omega$，它是所有有序的二元组的集合

$$\Omega_2 = \{\mathrm{ss}, \mathrm{sf}, \mathrm{fs}, \mathrm{ff}\}$$

$\mathscr{F}$包含有$2^4=16$个事件。比如ϕ，Ω，$\{\mathbf{ss}\}$，$\{\mathbf{ss}, \mathbf{ff}\}$，等等。

对于一般的n次伯努利试验，笛卡儿积样本空间变成了

$$\Omega_n = \underbrace{\Omega\times\Omega\times\cdots\times\Omega}_{n\text{次}}$$

包含有2^n个基本实验结果，每一个实验结果都是一个有序的n元组，因此

$$\Omega_n = \{a_1, \cdots, a_M\}$$

其中$M=2^n$，$a_i \triangleq z_{i_1}\cdots z_{i_n}$是一个有序的$n$元组，其中$z_{i_j}=s$或$f$，由于每一个实验结果$z_{i_j}$与其他实验结果是独立的，联合概率$P[z_{i_1}\cdots z_{i_n}]=P[z_{i_1}]P[z_{i_2}]\cdots P[z_{i_n}]$，因此，给定$k$个成功、$n-k$个失败的有序集的概率为$p^k q^{n-k}$。

例 1.9-1　（硬币投掷的重复试验）假定投掷硬币三次，每次投掷出现正面或反面的概率分别为$p=P(\mathrm{H})$和$q=P(\mathrm{T})$，事件$\{\mathrm{HTH}\}$的概率是$pqp=p^2q$。事件$\{\mathrm{THH}\}$的概率也是p^2q。下面列出了两次正面、一次反面的不同事件：

$$E_1 = \{\mathrm{HHT}\}$$

$$E_2 = \{\mathrm{HTH}\}$$

$$E_3 = \{\mathrm{THH}\}$$

如果F表示得到两次正面一次反面、且不考虑顺序这个事件，那么，$F=E_1\cup E_2\cup E_3$。由于对所有的$i\neq j$，$E_iE_j=\phi$，所以，$P[F]=P[E_1]+P[E_2]+P[E_3]=3p^2q$。

将前面的结果推广到由n次伯努利试验构成的实验，样本空间Ω_n包含$M=2^n$个实验结果a_1，a_2，…，a_M，其中每个a_i是n个符号的字符串，每一个符号表示成功或失败。考虑事件$A_k\triangleq\{$在n次试验中有k次成功$\}$，令带一撇的实验结果a_i'表示k次成功、$n-k$次失败的字符串，那么，用K表示在包含有k次成功、$n-k$失败的有序排列数，可得

$$A_k = \bigcup_{i=1}^{K}\{a_i'\}$$

为了确定K的大小，利用类似于在证明式(1.8-6)中用到的技巧。这里，用竖线表示失败，用星号表示成功，例如

$$|\ *\ *\ |\ *\ *\ |\ |\ *$$

表述在 9 次试验中 5 次成功，顺序为 fssfssffs。有多少次这样的排列呢？答案可由式(1.8-6)中用$r=k$和$(n-k)+k=n$代替$(n-1)+r$给出（注意，没有限制第一个符号和最后一个符号必须是竖线），因此

$$K = \binom{n}{k}$$

由于$\{a_i\}$是互不相交的，即对所有的$i \neq j$，$\{a_i\} \cap \{a_j\} = \phi$，由此可得

$$P[A_k] = P\left[\bigcup_{i=1}^{K}\{a_i'\}\right] = \sum_{i=1}^{K} P[a_i']$$

最后，由于无论 s 和 f 是什么顺序，$P[a_i'] = p^k q^{n-k}$，于是

$$P[A_k] = \binom{n}{k} p^k q^{n-k} \triangleq b(k; n, p) \tag{1.9-1}$$

二项式概率分布　式(1.9-1)定义的三参数函数$b(k; n, p)$称为二项式概率分布律，它是在n次独立的伯努利试验中k次成功的概率，其中每次伯努利试验成功的概率为p，二项式系数

$$C_k^n = \binom{n}{k}$$

在前一节就引入了，它是从容量为n的总体中形成容量为k的子总体数。在例 1.9-1 中，投掷硬币三次，总体容量为3(3 次试验)，子总体容量为2(两次正面)，我们感兴趣的是在三次试验中得到两次正面，而不考虑出现正面的顺序。因此，正确的结果是$C_2^3 = 3$。注意，如果要求前两次投掷中出现正面紧跟一次反面的概率，即$P[E_1]$，那么，就不能用系数C_2^3，因为这一事件的发生只有一种方式。

例 1.9-2　(从罐中抽出 2 个球)假定$n = 4$，即 4 个标有数字(1 ~ 4)的球，对于无重复的抽样，容量为 2 的可辨识的有序样本数是 12，即{1, 2}；{1,3}；{1,4}；{2,1}；{2,3}；{2, 4}；{3,1}；{3,2}；{3,4}；{4,1}；{4,2}；{4, 3}，可辨识的无序集合是数字 6，即

{1, 2} {2, 1}　　{1, 3} {3, 1}　　{1, 4} {4, 1}　　{2, 4} {4, 2}　　{3, 4} {4, 3}　　{2, 3} {3, 2}

由式(1.8-2)可直接得到这一结果，即($n = 4$，$k = 2$)

$$\binom{n}{k} = \frac{4!}{2!2!} = 6$$

例 1.9-3　(二元脉冲)每秒有 10 个独立的二元脉冲到达接收机。错误(即“0”接收为“1”，“1”接收为“0”)概率是 0.001，每秒至少有一个错误的概率为多少？

解

$$P[\text{至少有一个错误/秒}] = 1 - P[\text{没有错误/秒}]$$
$$= 1 - \binom{10}{0}(0.001)^0(0.999)^{10} = 1 - (0.999)^{10} \approx 0.01$$

思考　注意到

$$\sum_{k=0}^{n} b(k; n, p) = 1 \qquad \text{为什么？}$$

例 1.9-4　（奇数人出局）有一组奇数的人，他们要组织两支球队进行一场比赛，每支球队的人数是偶数，为了决定哪个人出局担任裁判，N 个人中的每一个按下列规定投掷均匀硬币，如果有一个人的结果（正面或反面）不同于该组剩余的人的结果，此人将会是裁判，假定有 11 个人，那么，在第一场比赛中，将成为“奇数人出局”人的概率是多少？

解　令 $E \triangleq \{10\text{H}, 1\text{T}\}$，其中 10H 意味着 H，H，…，H 十次，$F \triangleq \{10\text{T}, 1\text{H}\}$，那么，$EF = \phi$，且

$$\begin{aligned} P[E \cup F] &= P[E] + P[F] \\ &= \binom{11}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right) + \binom{11}{1}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{10} \\ &\approx 0.010\,74 \end{aligned}$$

例 1.9-5　（继续奇数人出局）在例 1.9-4 中推导了奇数人在第 n 次游戏第一次出局的概率的计算公式（提示：考虑将每次游戏看成独立的伯努利试验，如果奇数人出局发生为成功，否则为失败）。

解　令 E 表示在第 n 次游戏奇数人出局第一次发生的事件，令 F 表示在 $n-1$ 次游戏中，没有奇数人出局的事件，令 G 表示在第 n 次游戏奇数人出局的事件，那么

$$E = FG$$

由于假定 F 和 G 独立是完全合理的，因此，可以写成

$$\begin{aligned} P[E] &= P[F]P[G] \\ P[F] &= \binom{n-1}{0}(0.0107)^0(0.9893)^{n-1} = (0.9893)^{n-1} \\ P[G] &= 0.0107 \end{aligned}$$

因此，$P[E] = (0.0107)(0.9893)^{n-1}$，$n \geqslant 1$，常常称其为几何分布①。

例 1.9-6　（多彩票）如果要购买 50 张彩票，是应该在一种彩票中购买 50 张呢？还是在 50 种类似的彩票中各买一张？为简单起见，我们取 100 张票的情况，每张票的价格 1 美元，有 50 种独立的彩票可供选择。首先，考虑从这样一种彩票中购买 50 张。令 E_i 表示第 i 张彩票是一张中奖彩票，由于任何一张彩票都有可能是中奖彩票，且一张票顶多只有一个中奖人，由经典概率可得，$P[E_i] = n_{\text{win}}/n_{\text{tot}} = 1/100$，$l = 1, \cdots, 100$。赢得彩票这个事件是购买的 50 张彩票中有一张是中奖彩票，即等效于用 E 表示 50 张彩票中有一张是中奖的，$E = \cup_{i-1}^{50} E_i$ 和 $P[E] = P[\cup_{i-1}^{50} E_i] = \sum_{i=1}^{50} P[E_i] = 50 \times 1/100 = 0.5$。其次，考虑在 50 种不同彩票中各买一张这种情况。我们把它看成单次成功概率为 $p = 0.01$ 和 $q = 0.99$ 的伯努利试验。借助于计算器，可以求出一次中奖的概率为

$$\begin{aligned} P[\text{一次中奖}] &= b(1; 50, 0.01) \\ &= \binom{50}{1}(0.01)^1(0.99)^{49} \\ &= 50 \times 10^{-2} \times 0.611 \\ &\doteq 0.306^{②} \end{aligned}$$

---

① 关于这一定义的流行的变形是另一几何分布，即 pq^n，$n \geqslant 0$，其中 $q = 1 - p$，$0 \leqslant p \leqslant 1$。

② 等号上加一点表示小数点的数字是近似的。

类似地，两次中奖的概率为 $b(2;50,0.01)\approx 0.076$，三次中奖的概率为 $b(3;50,0.01)\approx 0.012$，四次中奖概率为 $b(4;50,0.01)\approx 0.001$，更多次中奖的概率就可以忽略不计了。我们很容易计算出至少中奖一次的概率为

$$\begin{aligned} P[\text{至少中奖一次}] &= 1-P[\text{每次都失败}] \\ &= 1-q^{50} \\ &= 1-(0.99)^{50} \\ &\doteq 0.395 \end{aligned}$$

实际上，我们有 0.395 =0.306 +0.076 +0.012 +0.001。因此，如果你只关心至少一次中奖的话，最好是从一种彩票中购买所有的 50 张。另一方面，当你在不同彩票中购买时，就有多次中奖的可能。所以更需要关注平均赢得的奖金。假定一个公平的彩票支付 100 美元，我们可以计算出平均赢得的奖金是

$$\begin{aligned} &100\times 0.306+200\times 0.076+300\times 0.012+400\times 0.001 \\ &\doteq 49.8 \end{aligned}$$

所以，根据平均赢得的奖金，这是一个相同的方法。

二项式概率分布的进一步讨论 先列出一些公式以便进一步使用。在 n 次试验中成功小于等于 k 次的概率为

$$B(k;n,p)=\sum_{i=0}^{k} b(i;n,p)=\sum_{i=0}^{k}\binom{n}{i}p^i q^{n-i} \tag{1.9-2}$$

符号 $B(k;n,p)$ 称为二项式分布函数，在 n 次试验中，成功大于等于 k 次的概率为

$$\sum_{i=k}^{n} b(i;n,p)=1-B(k-1;n,p)$$

成功大于等于 k 次但小于 j 次的概率

$$\sum_{i=k+1}^{j} b(i;n,p)=B(j;n,p)-B(k;n,p)$$

在后面的章中，有更多分布函数的讨论，例 1.9-7 给出一个这些公式应用的例子。

例 1.9-7 (导弹攻击)5 发导弹发射攻击海上的航空母舰，至少有 2 发命中的话将使航母沉没，所有 5 发导弹都处于正确的轨道上，但必须通过航母的“点防御”枪的拦截，我们知道“点防御”枪能够以概率 $p=0.9$ 摧毁导弹，在航母遇到导弹攻击后仍能飘浮在水面的概率是多少?

解 令 E 表示航母仍漂浮在水面的事件，F 表示导弹穿过“点防御”枪的事件，那么

$$P[F]=0.1$$

$$\begin{aligned} P[E] &= 1-P[E^c] \\ &= 1-\sum_{i=2}^{5}\binom{5}{i}(0.1)^i(0.9)^{5-i}\approx 0.92 \end{aligned}$$

多项式概率分布

多项式概率分布是二项式概率分布的推广，二项式概率分布是基于只有两个实验结果的

伯努利试验，而多项式概率分布是将伯努利试验推广到多 l 个可能的实验结果。因此，考虑一个基本实验，它是由一个具有 k 个可能的实验结果 ζ_1，ζ_2，…，ζ_l 的单一试验构成，令实验结果 ζ_i 的概率是 $p_i(i=1,\cdots,l)$，那么

$$p_i \geqslant 0, \qquad \sum_{i=1}^{l} p_i = 1 \tag{1.9-3}$$

假定将这一推广的伯努利试验重复 n 次，考虑一个预先规定的由基本实验结果的有序字符串构成的事件，在有序字符串中，ζ_1 出现 r_1 次，ζ_2 出现 r_2 次，以此类推，直到 ζ_l 出现 r_l 次。这个事件的概率是多少？这里的关键是预先规定的顺序。例如，$l=3$（三个可能的实验结果），$n=6$（6 次试验），预先规定的字符串是 $\zeta_1\zeta_3\zeta_2\zeta_2\zeta_1\zeta_2$，即 $r_1=2$，$r_2=3$，$r_3=1$，注意到 $\sum_{i=1}^{l} r_i = n$，由于每次试验的结果是独立的，预先规定的有序字符串的概率是 $p_1^{r_1}p_2^{r_2}\cdots p_l^{r_l}$，因此，字符串的 $\zeta_1\zeta_3\zeta_2\zeta_2\zeta_1\zeta_2$ 的概率是 $p_1^2p_2^3p_3$。

如果顺序没有指定，得出的概率也不同。假定将推广的伯努利试验重复 n 次，考虑 ζ_1 出现 r_1 次，ζ_2 出现 r_2 次，等等，不考虑顺序。在计算这个事件的概率之前，先看一个例子。

例 1.9-8 （繁忙的紧急号码）为了报告一个急诊去呼叫“救你生命”健康服务机构，有三件事之一可能会发生：

(1) 占线（事件 E_1）
(2) 号码错了（事件 E_2）
(3) 接通了值班护士（事件 E_3）

假定 $P[E_i]=p_i$，在不同时间 5 次不同的急诊中，初始呼叫遇到 4 次忙音，1 次拨错号码的概率是多少？

解 令 F 表示 4 次遇到忙音、1 次拨错号码的事件，那么

$$F = F_1 \cup F_2 \cup F_3 \cup F_4 \cup F_5, \quad \text{其中} F_1 = \{E_1E_1E_1E_1E_2\}$$

$$F_2 = \{E_1E_1E_1E_2E_1\},\ F_3 = \{E_1E_1E_2E_1E_1\},\ F_4 = \{E_1E_2E_1E_1E_1\}$$

和

$$F_5 = \{E_2E_1E_1E_1E_1\}$$

由于 $F_iF_j=\phi$，$P[F]=\sum_{i=1}^{5}P[F_i]$，而 $P[F_i]=p_1^4p_2^1p_3^0$ 独立于 i，因此

$$P[F] = 5p_1^4p_2^1p_3^0$$

假定 $p_1=0.3$，$p_2=0.1$，$p_3=0.6$，则

$$P[F] = 5 \times 8.1 \times 10^{-3} \times 0.1 \times 1 = 0.004$$

在这类问题中，我们必须对长度为 n 的所有字符串计数，ζ_1 出现 r_1 次，ζ_2 出现 r_2 次，等等。在前面的例子中，有 5 个这样的字符串。一般情况下，n 次试验中，实验结果 ζ_1 出现 r_1 次，实验结果 ζ_2 出现 r_2 次，等等，有

$$\frac{n!}{r_1!r_2!\cdots r_l!} \tag{1.9-4}$$

这样的字符串。在例 1.9-8 中，$n=5$，$r_1=4$，$r_2=1$，$r_3=0$，所以

$$\frac{5!}{4!1!0!} = 5$$

式(1.9-4)称为多项式系数。为了证明这个系数，考虑 r_1 个实验结果 ζ_1，在 n 次试验中，放置 r_1 个实验结果 ζ_1 的方法数与容量为 r 的总体中容量为 r_1 的子总体的个数 $\binom{n}{r_1}$ 相同，剩下的 $n-r_1$ 次试验，我们去放置 r_2 个实验结果 ζ_2，这样的方法数是 $\binom{n-r_1}{r_2}$，重复这一过程，得到的总的可辨识排列是

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\cdots\binom{n-r_1-r_2\ldots-r_{l-1}}{r_l}=\frac{n!}{r_1!r_2!\cdots r_l!}$$

例 1.9-9　(重复推广的伯努利试验)考虑 4 次重复的推广的伯努利实验，实验结果是 *，·，0，得到两个 *，一个 · 和一个 0 的总数是多少？

解　在 4 次试验中，得到两个 * 的总数是 $\binom{4}{2}=6$，如果把两个竖线之间的空格表示 1 次试验，那么，我们可以表示实验结果为

|*|*| | |; |*| |*| |; | |*|*| |; | |*| |*|; | | |*|*|; |*| | |*|

在两个剩余的单元中放置 · 的方法数是 $\binom{2}{1}=2$，在剩下的一个单元中放置 0 的方法数是 $\binom{1}{1}=1$，因此，总的排列是 $6\cdot2\cdot1=12$，它们是

*	0	•	*
*	•	0	*
*	*	0	•
*	*	•	0
*	0	*	•
*	•	*	0
0	*	*	•
•	*	*	0
0	*	•	*
•	*	0	*
0	•	*	*
•	0	*	*

下面讨论多项式概率分布，考虑一个推广的伯努利试验，实验结果为 $\zeta_1,\zeta_2,\cdots,\zeta_l$，令观测到 ζ_i 的概率为 p_i，$i=1,\cdots,l$，其中 $P_i\geqslant0$ 和 $\sum_{i=1}^{l}p_i=1$，在 n 次试验中 ζ_1 发生 r_1 次、ζ_2 发生 r_2 次，等等，这一事件发生的概率为

$$P(\boldsymbol{r};n,\boldsymbol{p})=\frac{n!}{r_1!r_2!\cdots r_l!}p_1^{r_1}p_2^{r_2}\cdots p_l^{r_l}\tag{1.9-5}$$

其中 $\boldsymbol{r}$ 和 $\boldsymbol{p}$ 是 l 元组，它们的定义为

$$\boldsymbol{r}=(r_1,r_2,\cdots,r_l),\quad \boldsymbol{p}=(p_1,p_2,\cdots,p_l),\quad \sum_{i=1}^{l}r_i=n$$

注意：令 $l=2$，式(1.9-5)变成了二项式概率分布，其中 $p_1\triangleq p$，$p_2\triangleq1-p$，$r_1\triangleq k$ 和 $r_2\triangleq n-k$。像式(1.9-1)和式(1.9-5)这样的函数常称为概率质量函数。

例 1.9-10　（紧急呼叫）在美国，911 是一个呼叫救护车、警察或消防部门的号码，在纽约州北部的喧嚣城市尼尔瓦纳，人们发现 60% 的 911 是呼叫警察，25% 呼叫救护车，15% 呼叫消防部门，我们观察下 10 个呼叫，6 次呼叫警察、3 次呼叫救护车和 1 次呼叫消防部门这样一个组合事件的概率是多少？

解　利用式(1.9-5)，可得到

$$P(6,3,1;10,0.6,0.25,0.15)$$
$$=\frac{10!}{6!3!1!}(0.6)^6(0.25)^3(0.15)^1\approx 0.092$$

如果 n 变得很大，会出现一个数值问题。例如，假定考察 100 次呼叫，认为 60 次呼叫警察、30 次呼叫救护车、10 次呼叫消防部门，显然需要计算 100!，60! 和 30! 这样的数，这时需要注意，帮助我们计算这样大的阶乘的一个很重要的结果就是斯特林(Stirling)①公式

$$n!\approx(2\pi)^{1/2}n^{n+(1/2)}\mathrm{e}^{-n}$$

其中近似随着 n 的增加而改善，例如

n	$n!$	斯特林公式	误差百分率
1	1	0.922137	8
10	3628800	3598700	0.8

当使用计算机计算式(1.9-5)时，要注意由于溢出而损失精度，使用对数对大的和小的数字是有帮助的。

正如前面所述，二项式概率分布是多项式概率分布的一种特殊情况，也是最重要的一种情况，当二项式概率分布的参数达到极端值时，二项式分布能够用来产生另一种重要的分布，这将在下面进行探讨。

1.10　二项式概率分布的渐近特性：泊松分布

假定在二项式函数 $b(k;n,p)$ 中，$n\gg 1$，$p\ll 1$，但是 np 保持一个常数，如 $np=\mu$，$q=1-p$，因此

$$\binom{n}{k}p^k(1-p)^{n-k}\approx\frac{1}{k!}\mu^k\left(1-\frac{\mu}{n}\right)^{n-k}$$

其中，如果 n 允许变得足够大，而 k 保持固定，则 $n(n-1)\cdots(n-k+1)\approx n^k$，因此，在 $n\to\infty$，$p\to 0$ 和 $k\ll n$ 的极限情况下，可得到

$$b(k;n,p)\approx\frac{1}{k!}\mu^k\left(1-\frac{\mu}{n}\right)^{n-k}\xrightarrow{n\to\infty}\frac{\mu^k}{k!}\mathrm{e}^{-\mu}\tag{1.10-1}$$

这样，二项式概率分布在应用到 $n\gg 1$，$p\ll 1$ 而 $np=\mu$ 是一个有限的常数的情况时

$$b(k;n,p)\approx\frac{\mu^k}{k!}\mathrm{e}^{-\mu}\tag{1.10-2}$$

泊松定律　参数为 μ 的泊松概率分布为

① James Stirling，18 世纪的数学家。

$$p(k)=\frac{\mu^k}{k!}\mathrm{e}^{-\mu},\quad 0\leqslant k<\infty$$

不像二项式概率分布，泊松概率分布刚好有一个可取任何正值的参数 μ。

例 1.10-1 (故障时间)一台计算机有 10 000 个器件，每个器件的失效是独立于其他器件的，且每个器件每年失效概率是 10^{-4}，计算机开机后连续工作一年的概率是多少？假定如果一个或多个器件失效计算机就不能工作。

解

$$p=10^{-4},\qquad n=10\,000,\qquad k=0,\qquad np=1$$

因此

$$b(0;10\,000\ \ 10^{-4})=\frac{1^0}{0!}\mathrm{e}^{-1}=\frac{1}{\mathrm{e}}=0.368$$

例 1.10-2 (随机点的时间)假定 n 个独立的点随机放置在 $(0,\ T)$ 上，令 $0<t_1<t_2<T$ 和 $t_2-t_1\triangleq\tau$，令 $\tau/T\ll1$ 和 $n\gg1$，在 τ 秒中精确地观测到 k 个点的概率是多少(参见图 1.10-1)？

解 考虑单个点随机放置在 $(0,\ T)$ 上，点出现在 τ 的概率为 τ/T。令 $p=\tau/T$，每个其他的点在 τ 秒出现的概率也相同，因此，在 τ 秒内发现 k 个点的概率是二项式概率分布

$$P[\text{在 }\tau\text{ 秒有 }k\text{ 个点}]=\binom{n}{k}p^kq^{n-k}\tag{1.10-3}$$

当 $n\gg1$ 时，利用近似式(1.10-1)，得

$$b(k;n,p)\approx\left(\frac{n\tau}{T}\right)^k\frac{\mathrm{e}^{-(n\tau/T)}}{k!}\tag{1.10-4}$$

其中 n/T 可以解释为每单位区间上的"平均"点数。

图 1.10-1 将点随机放置在一条线上，每个点等可能地放置在线的任何地方

在本例中用参数 $\mu(\mu>0)$ 代替平均率，我们得到如下泊松分布

$$P[k\text{ 个点}]=\mathrm{e}^{-\mu}\frac{\mu^k}{k!}\tag{1.10-5}$$

其中 $k=0,1,2,\cdots$，$\mu\triangleq\lambda\tau$，其中 λ 是每个单位时间的平均点数①，τ 是区间 $(t,\ t+\tau]$ 的长度，在流逝的时间 τ 中有 k 个点的概率是

$$P(k;t,t+\tau)=\mathrm{e}^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}\tag{1.10-6}$$

对于泊松概率分布，我们也规定在不相连接的时间间隔中到达的点数构成独立事件。可以把这个看成继承了伯努利试验构成的集合的特性，伯努利试验总是独立的。

在式(1.10-6)中，假定 λ 是常数，不是 t 的函数。如果 λ 随 t 变化，可以用积分 $\int_t^{t+\tau}\lambda(u)\mathrm{d}u$ 代替 $\lambda\tau$，在区间 $(t,\ t+\tau]$ 上有 k 个点的概率为

① 这里术语点是一个基本术语，等价的术语可以是"到达"、"命中"和"出现"。

$$P(k;t,t+\tau)=\exp\left[-\int_{t}^{t+\tau}\lambda(u)\,\mathrm{d}u\right]\frac{1}{k!}\left[\int_{t}^{t+\tau}\lambda(u)\mathrm{d}u\right]^{k} \tag{1.10-7}$$

泊松概率分布 P[在 Δx 上有 k 个事件]、或者更一般的情况是 P[在 $(x, x+\Delta x)$ 上有 k 个事件]在工程和科学中有广泛应用，其中 x 可以是时间、体积、距离，等等，Δx 是与 x 有关的区间。下面列出了在各领域中服从泊松概率分布的典型情况：

物理学：在放射性衰变中，P[在 τ 秒中的 k 个 α 粒子] 服从参数为 λ 的泊松分布，λ 代表每秒发射 α 粒子的平均数。

工程：在设计一个呼叫中心的容量时，P[在 τ 秒中的电话呼叫次数 k] 服从参数为 λ 的泊松分布，其中 λ 代表每秒的平均呼叫次数。

生物医学：在水污染监控中，P[在 1000 cm^3 中的大肠杆菌细菌数 k]服从参数为 λ 的泊松分布，其中 λ 为每立方厘米中的平均大肠杆菌细菌数。

交通：在设计高速公路收费站的容量时，P[τ 分钟内到达的汽车数 k]服从参数为 λ 的泊松分布，其中 λ 为每分钟的平均汽车数。

光学：在设计光子接收机时，P[在表面 A 上每秒光子数 k]服从参数为 λ 的泊松分布，其中 λ 为每单位面积上每秒的平均光子数。

通信：在设计光纤的光发射机-接收机链路时，P[1 s 内接收机产生的光电子数 k]服从参数为 λ 的泊松分布，其中 λ 为每秒的平均光子数。

参数 λ 也常常称为泊松率参数，它的度量是每个单位区间上的点，区间可以是时间、距离、体积，等等。当我们希望采用的泊松概率分布的形式是式(1.10-6)和式(1.10-7)时，我们说是服从率参数 λ 或率函数为 $\lambda(t)$ 的泊松分布。

例 1.10-3　(概率的误用)(a)用下面的数据证明地球外的宇宙肯定有生命①，平均每一个星系的恒星数为 300×10^9，而星系有 100×10^9 个，每一个恒星有一个行星系统的概率为 0.5，平均每个行星系统的行星数是 9 颗，行星能维持生命的概率为 1/9，在能维持生命的行星上出现生命的概率是 10^{-12}。

解　首先计算能维持生命的行星数量 n_{LS}

$$\begin{aligned} n_{\mathrm{LS}} &= 300\times10^9\times100\times10^9\times0.5\times9\times1/9 \\ &= 1.5\times10^{22} \end{aligned}$$

接着，用 $a=n_{\mathrm{LS}}p=1.5\times10^{22}\times10^{-12}$ 时二项式的泊松近似计算地球外无生命的概率，得

$$b(0,1.5\times10^{22},10^{-12})=\frac{(1.5\times10^{10})^0}{0!}\mathrm{e}^{-1.5\times10^{10}}\approx0$$

因此，我们刚好证明了地球外有生命的概率为 1，即这是一个肯定的事情。注意，其他行星上出现生命的概率是 10^{-12}，非常低的数字。

(b)现在除了将维持生命的星球出现生命的概率降到 10^{-30}，其他使用同样的数字，证明地球外有生命是极不可能的事情。

解　用 $a=1.5\times10^{22}\times10^{-30}=1.5\times10^{-8}$ 时二项式的泊松近似，得到地球外无生命的概率为

① 所有数据引自不同时期外星生命支持者提供的数据。

$$b(0,1.5\times10^{22},10^{-30})=\frac{(1.5\times10^{-8})^0}{0!}\mathrm{e}^{-1.5\times10^{-8}}$$
$$\approx 1-(1.5\times10^{-8})\approx 1$$

其中，我们利用了近似关系，对于小的 x，$\mathrm{e}^{-x}\approx 1-x$。

于是，可看到，仅仅改了一个数字，就从证明宇宙包含有地球外生命，变成了证明除了我们自己外宇宙是无生命的。概率误用的原因在于，当前我们并不了解从无生命的物质导致生命出现的因素是什么。当计算从技术讲是正确的时候，这个例子说明：当使用人为的数字去证明或不证明什么的时候，本质上是看你信仰或相信什么。

例 1.10-4 （网站服务器）一个网站服务器平均每分钟收到 16 个存取请求。如果服务器每分钟最多能够处理 24 个存取，那么，在任何一分钟内网站饱和的概率是多少呢？

解 如果一分钟内的请求数超过 24 次就出现饱和，这个事件的概率是

$$P[\text{饱和}]=\sum_{k=25}^{\infty}[\lambda\tau]^k\frac{\mathrm{e}^{-\lambda\tau}}{k!} \tag{1.10-8}$$

$$=\sum_{k=25}^{\infty}[16]^k\frac{\mathrm{e}^{-16}}{k!}\approx 0.017\approx 1/60 \tag{1.10-9}$$

因此，大约每 60 分钟会有一个访问者被拒绝。

知道了泊松分布在工程和科学中的许多应用后，人们或许会认为泊松分布的“高贵出生”只是二项式分布的极限。确实是这样，一旦给出三个假定，泊松分布就可以推导出来。显然，这三个假定只是物理过程特性的一种合理的描述。幸运的是，这些假定在许多情况下似乎是相当有效的。

为了更具体一点，我们将讨论在时间(或者长度、距离)上发生的事情。泊松分布基于下面三个假定：

1. 在 $(t,\ t+\Delta t]$ 中一个事件发生的概率正比于 Δt，即
$$P(1;t,t+\Delta t)\approx\lambda(t)\Delta t,\qquad \Delta t\to 0 \tag{1.10-10}$$
在式(1.10-10)中，$\lambda(t)$ 是泊松率参数。
2. 在 $(t,\ t+\Delta t]$ 中出现 $k(k>1)$ 个事件的概率趋于零
$$P(k;t,t+\Delta t)\approx 0\qquad \Delta t\to 0,\qquad k=2,3,\cdots \tag{1.10-11}$$
3. 在不重叠的时间区间上的事件是统计独立的①。

从这三个简单的物理假设开始，要得到泊松概率分布是一个简单的事情了。我们把推导留到第 9 章，这里只是指出，这些假设的运用导出基本的结果，即一阶差分方程，它的解是泊松分布。式(1.10-7)给出了一般的解，但幸运的是，在许多实际情况下，泊松率参数 $\lambda(t)$ 能够用一个常数 λ 来近似，这种情况下，式(1.10-6)可以应用。下面用最后一个例子来总结这一节的内容。

例 1.10-5 （数字磁带的瑕疵）计算机磁带的制造商发现，沿磁带长度方向的瑕疵密度是不均匀的，在对数据的仔细编辑后发现，长度为 D 的一条磁带，瑕疵密度 $\lambda(x)$ 沿磁带长度 x

---

① 注意，在性质 3 中我们谈论的是互不重叠的时间间隔，而没有交集的事件(互斥事件)，对于没有交集的事件，概率应该相加。而对于互不重叠的时间间隔，将导致泊松律的独立事件，概率应相乘。

的变化为

$$\lambda(x)=\lambda_0+\frac{1}{2}(\lambda_1-\lambda_0)\left(1+\cos\left[\frac{2\pi x}{D}\right]\right),\qquad \lambda_1>\lambda_0$$

其中 $0\leqslant x\leqslant D$，在边缘 $x=0$ 和 $x=D$ 磁带的瑕疵更大一些。

(a) 在这种情况下，$\lambda(x)$是什么意思?

(b) 长度为 D 的磁带的瑕疵平均数是多少?

(c) 长度为 D 的磁带的瑕疵 k 的表达式是什么?

(d) 在本例中的泊松假设是什么?

解

(a) 记住，$\lambda(x)$是瑕疵密度，即 x 处的每单位长度上的平均瑕疵数，因此我们得出结论，$\lambda(x)\Delta x$ 是磁带从 x 到 $x+\Delta x$ 的平均瑕疵数。

(b) 根据 $\lambda(x)$的定义，可得出结论，沿整个磁带的平均瑕疵数只是 $\lambda(x)$的积分，即

$$\begin{aligned}\int_0^D\lambda(x)\mathrm{d}x&=\int_0^D\left[\lambda_0+\frac{1}{2}(\lambda_1-\lambda_0)\left(1+\cos\frac{2\pi x}{D}\right)\right]\mathrm{d}x\\&=\frac{\lambda_0+\lambda_1}{2}D\\&\triangleq\Lambda\end{aligned}$$

(c) 假定泊松定律成立，利用式(1.10-7)，其中用 x 和 Δx(距离)分别代替 t 和 τ(时间)，因此

$$P(k;x,x+\Delta x)=\exp\left[-\int_x^{x+\Delta x}\lambda(\zeta)\mathrm{d}\zeta\right]\cdot\frac{1}{k!}\left[\int_x^{x+\Delta x}\lambda(\zeta)\mathrm{d}\zeta\right]^k$$

特别是在 $x=0$ 和 $x+\Delta x=D$ 时，可得到

$$P(k;0,D)=\Lambda^k\frac{\mathrm{e}^{-\Lambda}}{k!}$$

其中 Λ 在上面已经定义。

(d) 在本例使用了三个泊松假设:

(i) 当 $\Delta x\to 0$ 时，$P[1;x,x+\Delta x]\simeq\lambda(x)\Delta x$。

(ii) $P[k;x,x+\Delta x]=0$，$\Delta x\to 0$，$k=2,3,\cdots$，即在区间$(x,x+\Delta x)$上超过一个以上瑕疵的概率随 Δx 趋于零而变成零。

(iii) 在磁带非重叠部分出现瑕疵的事件是独立的。

1.11　二项式分布近似为正态分布

在本节，我们不加证明地给出二项式概率分布和二项式和的一个数值近似。令 S_k 表示在 n 次伯努利试验中由 k 次成功组成的一个事件，那么，k 的概率服从二项式分布

$$P[S_k]=\binom{n}{k}p^kq^{n-k}=b(k;n,p),\quad 0\leqslant k\leqslant n \tag{1.11-1}$$

对于大的 n 和 k 的值，式(1.11-1)可能难以数值计算。此外，事件$\{k_1<$ 成功数$\leqslant k_2\}$的概率可能包含许多项，直接计算概率 $P\{k_1<$ 成功数$\leqslant k_2\}$是很困难的。幸运的是，对于大的 n，我们能够用近似的方法计算这个概率。这些近似的方法包括了所谓的正态或高斯分布。

正态分布和它的意义将在第2章以及本书的随后几章中详细讨论，这里只是用它来帮助计算二项式概率，在这里我们定义一个函数$f_{SN}(x)$，称为标准正态分布

$$f_{SN}(x) \triangleq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \tag{1.11-2}$$

它的积分称为标准正态累积分布函数

$$F_{SN}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y \tag{1.11-3}$$

当n很大时，可以证明[1-8]

$$b(k;n,p) \approx \frac{1}{\sqrt{npq}} f_{SN}\left(\frac{k-np}{\sqrt{npq}}\right) \tag{1.11-4}$$

当$npq \gg 1$时，近似的效果变得更好。在表1.11-1中重新给出了近似的结果[1-8]，甚至在$npq = 1.6$这样的情况，近似的效果也是相当好的。当$n \gg 1$，k_1和k_2固定为整数，那么和的近似为

$$P[k_1 < \text{成功数} \leqslant k_2] \approx F_{SN}\left[\frac{k_2 - np + 0.5}{\sqrt{npq}}\right] - F_{SN}\left[\frac{k_1 - np - 0.5}{\sqrt{npq}}\right] \tag{1.11-5}$$

对于不同的值n，p，k_1，k_2，近似结果在表1.11-2给出，这些结果引自参考文献[1-8]。

表1.11-1 选定数后用正态分布去近似二项式分布

k	$b(k;10,0.2)$	正态近似
0	0.1074	0.0904
1	0.2864	0.2307
2	0.3020	0.3154
3	0.2013	0.2307
4	0.0880	0.0904
5	0.0264	0.0189
6	0.0055	0.0021

表1.11-2 用正态近似计算事件的概率(引自参考文献[1-8])

n	p	α	β	$P[\alpha \leqslant S_n \leqslant \beta]$	正态近似
200	0.5	95	105	0.5632	0.5633
500	0.1	50	55	0.3176	0.3235
100	0.3	12	14	0.000 15	0.000 33
100	0.3	27	29	0.2379	0.2341
100	0.3	49	51	0.000 05	0.000 03

在采用正态近似时，应该参考表2.4-1，在表2.4-1中，给出的函数为$\mathrm{erf}(x)$而不是$F_{SN}(x)$，$\mathrm{erf}(x)$定义为

$$\mathrm{erf}(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_{0}^{x} \mathrm{e}^{-\frac{y^2}{2}} \mathrm{d}y$$

然而，很容易证明

$$F_{SN}(x) = \tfrac{1}{2} + \mathrm{erf}(x), \quad x > 0 \tag{1.11-6}$$

和

$$F_{SN}(x) = \tfrac{1}{2} - \mathrm{erf}(|x|), \quad x < 0 \tag{1.11-7}$$

可以根据表的值计算式(1. 11-5)，因此，用 $a \triangleq \frac{k_1 - np - 0.5}{\sqrt{npq}}$ 和 $b \triangleq \frac{k_2 - np + 0.5}{\sqrt{npq}}$，$b' > a'$，我们就可以利用表 1. 11-3 的结果。

正态近似在计算泊松和时也是有用的。例如，像式(1. 10-9)这样的求和，如果直接计算是非常繁琐的。然而，如果 $\lambda\tau \gg 1$，就可以使用正态去近似泊松分布，它只是正态去近似二项式分布的简单扩展。这种近似是我们所期望的，因为我们已经看到，泊松分布自身就是在某种条件下对二项式分布的近似。从以上给出的结论，能够证明下列近似：

$$\sum_{k=\alpha}^{\beta} \mathrm{e}^{-\lambda\tau} \frac{[\lambda\tau]^k}{k!} = \frac{1}{\sqrt{2\pi}} \int_{l_1}^{l_2} \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y \tag{1.11-8}$$

其中

$$l_2 \triangleq \frac{\beta - \lambda\tau + 0.5}{\sqrt{\lambda\tau}}$$

和

$$l_1 \triangleq \frac{\alpha - \lambda\tau - 0.5}{\sqrt{\lambda\tau}}$$

另一个有用的近似为

$$\mathrm{e}^{-\lambda\tau} \frac{[\lambda\tau]^k}{k!} = \frac{1}{\sqrt{2\pi}} \int_{l_3}^{l_4} \exp\left(-\frac{1}{2}y^2\right) \mathrm{d}y \tag{1.11-9}$$

其中

$$l_4 \triangleq \frac{k - \lambda\tau + 0.5}{\sqrt{\lambda\tau}}$$

和

$$l_3 \triangleq \frac{k - \lambda\tau - 0.5}{\sqrt{\lambda\tau}}$$

例如，当 $\lambda\tau = 5$，$k = 5$，式(1. 11-9)用正态近似的误差小于 1% 。

小结

在本章，我们复习了概率的不同定义，给出了概率的公理化理论，并且说明了对于随机实验，要求有三个重要的目标：样本空间 Ω、事件的 $\mathscr{F}$ 域和概率测度 P。数学元组(Ω, $\mathscr{F}$, P)称为概率空间 $\mathscr{P}$。

我们介绍了一些重要的术语，如独立、非独立、复合事件、条件概率等。并且给出了一些能够应用的关系。

我们讨论了组合的一些公式，并且简要说明了在理论物理中的重要性。接着讨论了二项式概率分布及其他的推广——多项式分布。我们看到，二项式分布在某些极限条件可以用泊松分布近似，作为概率论中心极限定理之一，泊松定律在科学和工程中有着广泛的应用。泊松分布能够直接从简单合理的物理假设推导出来，详细的验证放将在第 9 章中讨论。

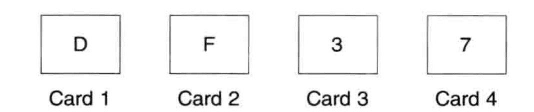
我们提供了用正态分布去近似二项式分布和泊松分布。在第 4 章中讨论工程中感兴起的入住问题时将再次用到这些近似关系。

习题

(＊星号的习题是一些较难的题目，可能需要多做些工作，或者需要附加的阅读。)

1.1 对于像“拉尔夫(Ralph)可能犯有盗窃罪”这样的语句，从相对频率的概率意义上需要什么样的数据?

1.2 在应用概率中(数学中一个分支，称为统计学)常常涉及到检验 $P \to Q$(P 意味着 Q)类型的语句，例如，如果她抽烟，那么她有可能得病;如果她聪明，她在学校就会表现得好。给你四张牌，每张牌在一边上有个字母，而另一边上有个数字，要求你测试一个规则“如果牌一边是 D，另一边是 3，你应该翻转下面牌中的哪一张来验证规则的真实性。



这里要仔细一点。

1.3 在一个转轮游戏中，转轮包含数字 1 ~9，如果转到偶数，参赛者就赢了，参赛者获胜的概率是多少?你的假设是什么?

1.4 一枚均匀硬币投掷三次，每次投掷的结果是正面 H 或反面 T，得到两次反面一次正面的概率是多少?

1.5 一个罐子装有三个球，分别标有数字 1、2、3，实验为随机抽取一个球，抽出的球在下次抽取球之前被放回原处，称为重复抽样，在三次实验中抽得到同一个球的概率是多少?

1.6 一个罐子装有 6 个球，分别标有数字 1 ~6，实验为随机抽取一个球，并且抽取后不放回原处，描述样本空间 Ω。如果抽取后放回原处，样本空间 Ω 又是什么?

1.7 实验为测量随机选择的一组夫妻的身高。(a)用简便的符号描述 Ω;(b)令 E 表示丈夫比妻子矮这样的事件，用简便的符号描述 E。

1.8 一个罐子装有 10 个球，分别标有数字 1 ~10，令 E 表示抽取一个不大于 5 的球，令 F 表示抽取一个大于 3 小于 9 的球。计算 E^c，F^c，EF，$E\cup F$，EF^c，$E^c\cup F^c$，$EF^c\cup E^cF$，$EF\cup E^cF^c$，$(E\cup F)^c$ 和 $(EF)^c$，用文字表示这些事件。

1.9 有四个等可能的实验结果 ζ_1、ζ_2、ζ_3 和 ζ_4，以及两个事件 $A=\{\zeta_1,\zeta_2\}$ 和 $B=\{\zeta_2,\zeta_3\}$，用它们的元素(实验结果)表示事件 AB^c，BA^c，AB 和 $A\cup B$。

1.10 验证有用的一组等式 $A=AB\cup AB^c$ 和 $A\cup B=(AB^c)\cup(BA^c)\cup(AB)$，这些并的概率如何计算?为什么?

1.11 在一个给定的随机实验中，假定有四个等可能的实验结果 ζ_1、ζ_2、ζ_3 和 ζ_4，令事件 $A\triangleq\{\zeta_1,\zeta_2\}$，$A$ 的概率是多少?事件(集合)A^c 用实验结果如何表示?验证 $P[A]=1-P[A^c]$。

1.12 考虑概率空间$(\Omega, \mathscr{F}, P)$，

(a)阐述概率的三个公理，每一个公理用合适的语言加以叙述。

(b)推导下列公式，用以上合适的公理证明推导的每一步骤

$$P[A\cup B]=P[A]+P[B]-P[A\cap B]$$

其中，A 和 B 是域 $\mathscr{F}$ 中的任意事件。

1.13 实验为从一个盛有 5 个球(分别标有数字 1 ~5)的罐子里随机抽取两个球，采用有重复抽样的方法，要求三个学生“Dim”，“Dense”和“Smart”去计算两次抽取球的和等于 5 的概率。Dim 求得概率为 $p=2/15$，他认为有 15 个不同的无序对，只有两对符合条件。即(1,4)和(2,3);Dense 求得概率为 $p=1/9$，他认为有 9 个不同的和(2 ~10)，只有一个符合条件。Smart 求得概率为 $p=4/25$，他认为有 25 个不同的有序结果，其中 4 个符合条件。即(4,1)，(3,2)，(2,3)和(4,1)。为什么 $p=4/25$ 是正确的答案呢?解释为什么 Dense 和 Dim 的推理是错误的?

1.14 证明交集的分配率，即

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

可以通过证明每一边包含在另一边中来达到目的。

1.15　对于任何概率实验和定义在这个实验上的事件 A，证明：$P[A]=1-P[A^c]$。

1.16　令 $\Omega=\{1,2,3,4,5,6\}$，定义三个事件：$A=\{1,2\}$，$B=\{2,3\}$ 和 $C=\{4,5,6\}$，概率测度未知，但是它满足三个公理。

(i) $A\cup C$ 的概率是多少？

(ii) $A\cup B\cup C$ 的概率是多少？

(iii) 叙述 B 和 C 是独立事件的条件。

1.17　用式(1.5-1)到式(1.5-3)给出的公理证明下列关系：$(E\in\mathscr{F},\ F\in\mathscr{F})$ (a) $P[\phi]=0$；(b) $P[EF^c]=P[E]-P[EF]$；(c) $P[E]=1-P[E^c]$。

1.18　对于本题，采用概率空间 $(\Omega,\mathscr{F},P)$，实验结果、事件和事件的域有什么区别？

1.19　利用的概率的公理化定义证明：$(A\in\mathscr{F},\ B\in\mathscr{F})$：$P[A\cup B]=P[A]+P[B]-P[A\cap B]$，其中 P 是样本空间，$\mathscr{F}$ 是事件的域。

1.20　利用式(1.4-3)证明：$P[E\oplus F]=P[EF^c]+P[E^cF]$。

1.21　证明前一题中的 $P[E\oplus F]$ 能够写成 $P[E\oplus F]=P[E]+P[F]-2P[EF]$。

1.22 *　令样本空间为 $\Omega=\{$猫,狗,山羊,猪$\}$

(a) 假定只给定下列概率信息：

$P[\{$猫,狗$\}]=0.9$

$P[\{$山羊,猪$\}]=0.1$

$P[\{$猪$\}]=0.05$

$P[\{$狗$\}]=0.5$

对于给定的这些概率，求一个合适的事件域 $\mathscr{F}$，以至于能定义出概率空间 $(\Omega,\mathscr{F},P)$，通过列出域的事件以及对应的概率来描述域。

(b) 如果没有 $P[\{$猪$\}]=0.05$ 的概率信息，重复(a)。

1.23　通过证明每一边包含在另一边中证明交集的分配律，即

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

1.24　在一个给定的随机实验中，有四个等可能实验结果 ζ_1、ζ_2、ζ_3 和 ζ_4。令事件 $A=\{\zeta_1,\zeta_2\}$，A 的概率是多少？用实验结果表示的 A^c 是多少？A^c 的概率是多少？验证 $P[A]=1-P[A^c]$。

1.25　一个罐子里装有 8 个球，用字母 a 和 b 来对球进行标注，两个球标上 a，两个标上 b，剩下的球标上字母 ab，除了标注不同外，所有球都是相同的。现在随机地从罐子里取球，令 A 和 B 分别表示观测到字母 a 和 b，求 $P[A]$、$P[B]$ 和 $P[AB]$。A 和 B 独立吗(注意，如果你抽出的球是 a 球或者 ab 球，那就意味着你观测到字母 a)？

1.26　一枚均匀骰子投掷两次(均匀骰子是指所有实验结果 1，…，6 是等可能的)。假定在第一次投掷出现 3，第二次投掷后和为 7 的概率是多少？

1.27　从一副 52 张的扑克牌中随机抽取一张牌，令 A 表示抽到"A"牌的事件、B 表示抽到红牌的事件，在一副扑克牌中有 4 张"A"、26 张红色的牌，A 和 B 独立吗？

1.28　一枚均匀的骰子投掷三次，假定在第一次投掷时出现 2，三次投掷得到和为 7 的概率是多少？

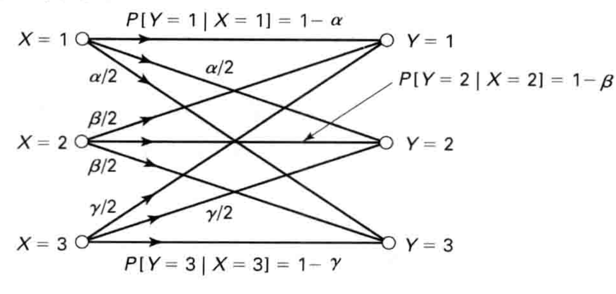
1.29　一随机数产生器产生整数 1 ~9(互斥)，所有的实验结果是等可能的，每个整数都是与任何前面的整数独立产生的。令 Σ 表示两个连续产生的整数和，即 $\Sigma=N_1+N_2$。假定 Σ 是奇数，Σ 为 7 的条件概率是多少？假定 $\Sigma>10$，至少有一个整数大于 7 的条件概率是多少？假定 $N_1>8$，Σ 为奇数的条件概率是多少？

1.30　著名的哈佛医学院(HMS)有 60 名学生和医生。假定有一个检测某种疾病 D 的测试，这种疾病的流行率为 0.001，即随机选择的一个人患上 D 的概率 $P[D]$ 为 0.001。测试的假阳性率 0.005，正确检测率是 1。所谓正确检测是指患有 D 的人被正确检测出来。假定你被检测到 D 为阳性，你实际患有 D 的概率

是多少？哈弗医学院的许多专家的答案是0.95，平均答案是0.56。如果你得出0.17的正确答案，那表明你的概率知识要比哈佛医学院专家们的概率知识多。

1.31　亨丽埃塔(Henrietta)已经29岁了，身体很健康。她在大学主修地质学。她在学生时代，经常到国家森林公园徒步旅游和在国家公园骑自行车。她参加了反伐木和反采矿的运动。现在亨丽埃塔在尼瓦纳(Nirvana)镇办公大楼里工作。亨丽埃塔的职业是办公室经理的概率或亨丽埃塔是一名像西拉(Sierra)俱乐部那样的自然保护组织办公室经理的概率，这两个概率哪个大？

1.32　在图P1.32的三元通信信道中，发送3是发送1的三倍，发送2是发送1的两倍。现在观察到1，那么发送的是1的条件概率是多少？



图P1.32　三元通信信道

1.33　一个大班正在进行概率论的考试，对于某个测试题，知道答案的考生比率是p，猜测的考生比率是$1-p$。知道答案的考试正确答题的概率是1，猜测的概率$1/m$，m是多项选择的备选数。计算在他或她正确答题时知道答案的概率。

1.34　在例1.6-12的选美问题中，如果我们预先决定在第$i(1\leqslant i\leqslant N)$次选出一个候选者，那么选到最美候选者的概率是多少？

1.35　假定在一个生产芯片的半导体制造厂有三台机器A、B、C，它们分别生产那个厂的芯片的25%、35%和40%。在所有生产出的产品中，分别有5%、4%和2%的芯片是次品。随机从三台机器输出的产品中抽取一个芯片，并且发现是次品，这个次品分别由机器A、B或C生产的概率是多大？

1.36　在例1.6-12中，假定采用“等等看”的策略，画出正确判决的概率与k/N的曲线，特别是当$k/N=0.5$时，$P[C]$是多少？当k离k_0不远、且N大时，$P[C]$对k的灵敏度有什么建议？

1.37　在圣·奥里伽米的马德里拉巴斯的一个村子里，一场大洪水使得103名村民无家可归，政府建了30个帐篷的临时帐篷村，将103名村民随机地安排到30个帐篷里。

(a)把此问题看作为一个入住问题，那么什么是与入住问题相似的球？什么又是单元？

(b)将村民分配到帐篷中，可以做多少种不同的分配？

(c)有多少种没有一个帐篷空着的不同分配？

1.38　考虑r个难分辨的球(粒子)和n个单元，其中$n>r$，r个球随机地放入n个单元(入住多个是可能的)，r个球出现在预先选定的r个单元(每个单元一个)的概率P是多少？

1.39　假定有r个难识别的球和n个单元，单元最多只能放一个球，且$r<n$，r个球出现在预先选定的r个单元的概率是多少？

1.40　三位部落长老赢得了瓦提斯坦北部的选举，作为苏达比亚人民的礼物，5把相同的突击步枪被空投到三人会面的会场，部落领导者拼命地奔跑去收集尽可能多的能带走的突击步枪。

(a)把此问题看成一个入住问题。

(b)列出所有可能的在三名部落领导者的不同分配。

(c)有多少种不同的分配，其中至少有一名部落领导者没有捡到任何突击步枪。

(d)三名部落领导者至少捡到一把突击步枪的概率是多少？

(e)只有一名部落领导者没有见到突击步枪的概率是多少？

1.41　在某些赌场有一种骰宝游戏，赌徒对投掷三粒骰子的结果下赌注。有多种下赌注的方法，每一种有不

同的赔付，我们列出下面一些用括号表示的赔付：

(a)指定的三个相同数字(1 赔 180)。

(b)未指定的三个相同数字(1 赔 30)。

(c)指定两个相同的数字(1 赔 10)。

(d)三个骰子之和等于 4 或 17(1 赔 60)。

(e)三个骰子之和等于 5 或 16(1 赔 30)。

(f)三个骰子之和等于 6 或 15(1 赔 17)。

(g)三个骰子之和等于 7 或 14(1 赔 12)。

(h)三个骰子之和等于 8 或 13(1 赔 8)。

(i)三个骰子之和等于 9、10、11、12(1 赔 6)。

(j)指定两个骰子的组合，即在三个显示的骰子中，必须有两个精确地匹配一个组合的下注(1 赔 5)。

我们希望从玩家的观点计算赢的概率以及他期待的收获。

1.42　大多数通信网络都采用分组交换在两个用户之间建立虚拟电路，即使用户与另一个用户共享同一物理信道。在分组交换中，数据流被分成若干包在不同的路径传输，并在正确的地址以合适的时间顺序重装，假定顺序信息丢失了，计算即使没有顺序信息，分成 N 个包的数据流能正确地重装的概率。

1.43　在前一题中假定 $N=3$，一位懒惰的工程师觉得省去顺序信息，而采用重复发送数据流直到数据包首次正确地重新排序为止，推导在第 n 次发送时第一次出现正确地重新排序的公式。第一次正确地重新排序的累积概率至少达到 0.95 之前，重复发生的次数应为多少？

1.44　通过证明 $\sum_{k=0}^{n} b(k;\ n,\ p) = 1$ 来证明二项式分布律 $b(k;n,p)$ 是一个有效的概率赋值。

1.45　考虑一个战争生存游戏。有 6 枚来袭的弹道导弹(BM)对 12 枚点火了的反弹道导弹(AMM)，两枚 AMM 对一枚 BM，AMM 的单次拦截概率(SSKP)是 0.8，SSKP 是 AMM 拦截 BM 的概率。假定 AMM 互不干扰，并且一枚 AMM 最多只能拦截一枚 BM，计算下列概率：(a)所有 BM 被拦截；(b)至少一枚 BM 突防成功并摧毁目标；(c)刚好一枚 BM 突防成功。

1.46　在前一题中假定目标被 BM 摧毁。计算只有一枚 BM 突防成功的条件概率？

1.47　计算机芯片制造商发现，每生产 100 个芯片，85 个满足技术指标，10 个需要重新生产，5 个为废品。现在挑选 10 个芯片进行检查：

(a)10 个均满足技术指标的概率？

(b)2 个或 2 个以上为废品的概率？

(c)8 个满足技术指标，1 个需要重新生产，1 个为废品的概率？

1.48　不像在纽约，911 是应急的通用电话号码，在俄罗斯的莫斯科，拨 01 是火警，02 是找警察，03 是急救。据估计在俄罗斯应急电话与纽约有大致相同的分布，60% 是找警察，25% 是急救服务，15% 是火警。假定监控到 10 次呼叫，这些呼叫在时间上不重叠，且这些呼叫构成了独立的试验：

(a)10 次拨号产生序列 02030202030102030202 的概率是多少？

(b)有多少个不同的序列包含有 6 次呼叫找警察、3 次呼叫找救护车、1 次找火警？

1.49　一个走私犯企图冒充玻璃珠进口商，试图把钻石混合在玻璃珠中来走私钻石，混合的比例是每 1000 个珠中有一颗钻石珠。海关检察员检查了 100 个样本，走私犯被逮住的概率是多少？即在抽样中至少有一颗钻石珠的概率。

1.50　假定一个容错接收机对听众大的烦恼声产生一种有声点击，每秒平均点击数依赖于接收机的温度，可表示为 $\lambda(\tau) = 1 - e^{-\tau/10}$，其中 τ 是开机时间，求开机后前 10 s 点击 0、1、2…的概率，假定开机后点击数服从泊松分布。

1.51　有一个频繁发行的彩票，每次发行时以 1 美元的价格发行 100 张，每次发行肯定有一个赢家。有一个玩家花了 50 美元，为了使每次发行中至少有一次中彩的概率最大，在一次发行中他应该买 50 张彩票吗？或者在 50 次彩票发行中他应该每次买一张吗？

1.52 在前一个习题中，两种策略的哪一种策略使得玩家有更大的收益？如果玩 $M(M\leqslant 50)$ 次彩票，期待的收益定义为 $\bar{G}_M \triangleq \sum_{i=1}^{M} G_i P(i)$，其中 G_i 是在第 i 次彩票发行时赢得的收益。

1.53 在一个如图 P1.53 的开关网络表示数字通信链路，开关 $\alpha_i(i=1, 2, \cdots, 6)$ 的开或者合，它们是独立操作的，开关合的概率是 p。令 A_i 表示开关 i 是合的事件。

(a) 用 A_i 表示从 1 ~2 至少存在一条合的路径。

(b) 计算从 1 ~2 至少存在一条合的路径的概率。

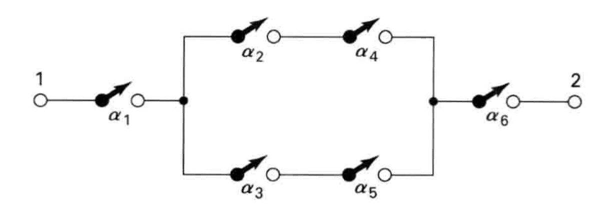


图 P1.53 电话链路中的开关

1.54 (在互不相交的区间上服从泊松分布的事件的独立性)每分钟到达公路收费站的平均车辆数为λ，在 $(0,T)$ 分钟间隔中有 k 辆车到达的概率为

$$P(k;0,T) = \mathrm{e}^{-\lambda T}\frac{[\lambda T]^k}{k!}$$

考虑两个不相交的区间 $(0, t_1]$ 和 $(t_1, T]$，即不重叠的区间，由泊松分布：

$$P[\text{在}(0, t_1]\text{上有 } n_1 \text{ 辆车和}(t_1, T]\text{上有 } n_2 \text{ 辆车}]$$
$$=P[\text{在}(0, t_1]\text{上有 } n_1 \text{ 辆车}]P[(t_1, T]\text{上有 } n_2 \text{ 辆车}]$$

即在不相交区间上的事件是独立的。利用这样的事实证明：

(a) $P[$在$(0, t_1]$上有 n_1 辆车|在$(0, T]$上有 n_1+n_2 辆车$]$与 λ 无关。

(b) 在(a)中令 $T=2$，$t_1=1$，$n_1=5$ 和 $n_2=5$，计算 $P\{$在$(0,1]$上有 5 辆车|在$(0,2]$上有 10 辆车$\}$。

1.55 在麻醉中使用的自动呼吸机(B)的故障概率为 P_B，故障意味着病人死亡，除非监视系统(M)检测到故障并提醒外科医生。监视系统故障的概率为 P_M，系统部件的故障是独立的事件。哈佛医学院的 X 教授(医学博士)认为,如果 $P_M > P_B$，M 的安装就是无用的①。证明 X 教授需要选修概率论的课程以便计算用和不用监视系统病人死亡的概率，取 $P_M=0.1=2P_B$。

1.56 在一个特定的通信网中，服务器传播一个数据包给 N 个接收机，然后服务器在规定的时间内等待从 N 个接收机来的确认消息，之后再传播下一个数据包。如果在规定的时间内没有接收到确认消息，服务器就重发同一个数据包。服务器被称为"重传模式"，它继续重发直到收到所有 N 个确认信息，然后再发下一个数据包。令 $p \triangleq P[$一个数据包成功发送到一个接收机并收到确认消息$]$。假定对不同的接收机所有这些事件是独立的，或者传送是分开的。由于在传输媒介中的随机损耗以及接收机的可变的条件，我们有 $p<1$。

(a) 在固定的协议或者操作方法中，我们要求接收到所有 N 个确认消息来对给定的传输尝试做出响应，以便宣告数据包传输成功。令事件 $S(m)$ 定义为 $S(m) \triangleq \{$数据包对所有 N 个接收机在小于等于 m 次尝试后成功传送$\}$。求概率

$$P(m) \triangleq P[S(m)]$$

[提示:考虑事件 $S(m)$ 的补。]

(b) 一个改进的系统是根据下面的动态协议来运行的:这里我们放松对重传输尝试的确认消息的要求，而只要求那些从前一次尝试到当前数据包发送还没有听到的那些接收机的确认消息。令 $S_D(m)$ 是与(a)中相同的事件，但使用了动态协议。求概率

① 这是一个真实的故事，医学院的名称已改变。

$$P_D(m) \triangleq P[S_D(m)]$$

[提示:首先考虑每个接收机事件 $S_D(m)$ 的概率，然后推广到 N 个接收机。]

注意:如果你用 $p=0.9$ 和 $N=5$ 来试试的话，你应该发现 $P(2)<P_D(2)$。

1.57　投掷两枚无偏差的骰子(每个骰子有 6 个面:1 ~6)，写出两个面的数字和。重复这个过程 100 次。有 10 次的数字和为 7 的概率是多少? 计算这个概率的泊松近似是什么(提示:考虑在单次投掷中 $A=\{和=7\}$ 这个事件，并且在式(1.9-1)中的 p 为 $P[A]$)?

1.58　为客户考虑，你必须提供洗衣设备，你的选择是:

1. 以每月 50 美元的价格租两台较为便宜的“Clogger”洗衣机;或者
2. 以每月 100 美元的价格租一台“NeverFail”洗衣机。

“Clogger”有 40% 的时间不能工作，而“NeverFail”有 20% 的时间不能工作。

(a)按客户的观点，哪一个选择更好?

(b)按照你是老板的角度考虑，哪一个选择更好?

1.59　在 Borduria 东部的政局不稳定的国家里，常常发现乘客携带炸弹上飞机，任意给定一个航班，飞机上有炸弹的概率是 10^{-2}，一位胆小的乘客总是在手提箱里带有一个废弃的炸弹，他认为有两颗炸弹在飞机上的概率是 10^{-4}。按这样的策略，胆小的乘客相信它已经大大地降低了飞机被炸毁的机会。你同意他的推断吗? 如果不同意，为什么?

1.60　在一个由 8 条链路组成的环形网络中(如图 P1.60 所示)，有两条路径连接到任意两个终端。假定链路独立失败的概率是 q，$0<q<1$，求数据包成功由终端 A 传送到 B 的概率(注意:终端 A 向环的两个方向发送数据包，终端 B 从环的接收方向移出数据包。成功的传送意味着终端 B 从两个方向接收到数据包)。

1.61　工会指示电话公司的行政部门，要求对 8 小时班中处理了超过 5760 个呼叫的电话接线员付加班费。加入了工会的电话接线员柯蒂斯在某天的 8 小时班中出现的呼叫服从率参数为 $\lambda=720$ 呼叫/小时的泊松分布，他收到了加班费的概率是多少?

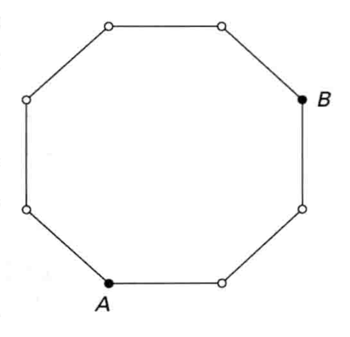


图 P1.60　8 个站的环形网络

1.62　投掷两枚无偏的硬币(每枚硬币有两面，分别标有 1 和 2)，记下两面数字之和，重复这个过程 80 次。得到 10 次和数为 2 的概率是多少? 计算这个概率的泊松近似是多少?

1.63　到达收费站的平均车辆数是每分钟 λ 辆，假定到达的车辆数服从泊松分布。假定在前 2 分钟到达了 5 辆车，在前 4 分钟到达 10 辆车的概率是多少?

1.64* 有一位老教授，急切希望学生对他的课程有一个好的评价，于是他给每个学生派发巧克力。教授的短期记忆是如此之差，以至于记不起哪个学生收到了巧克力。假定巧克力是随机派发的，有 10 个学生和 15 块巧克力，每个学生至少收到一块巧克力的概率是多少?

1.65　假定计算机代码错误发生如下:一行代码含有错误的概率是 $p=0.001$，无错误的概率是 $q=0.999$，不同行发生代码错误是独立的。在一个 1000 行的程序中，找到 2 个或更多错误行的概率是多少?

1.66　如果两个人的生日的月和日是相同的(年不必是相同的)，我们就认为这两个人是同一天生日。在屋里需要有多少人才能使得至少有两人的生日在同一天的概率等于或大于 1/2?

1.67　(抽样)我们从一条半导体生产线随机地抽出 10 个芯片，已知芯片的缺陷率为 2% ,求抽出的芯片中有缺陷的芯片大于 1 的概率。

1.68* (渗流分形)考虑一个方形格子，有 N^2 个单元，即每一边有 N 个单元。写一个程序做如下事情:以概率 p 将一个导电的元素放入单元中，以概率 $q=1-p$ 让单元空着，对格子里的每一个单元都这么做。当你做完后，存在一条路径让电流从格子的底部到顶部吗? 如果是的，就说格子是渗流的。渗流模型用在流行病、森林火灾、Ad hoc 网络等的研究中。格子也称为随机分形，因为它具有某种不变的性质。试试 $N=10$、20、50，$p=0.1$、0.3、0.6。你需要随机数产生器，MATLAB 有一个函数 rand，它可以产生在

(0.0,1.0)上均匀分布的随机数 x_i。如果随机数 $x_i \leq p$，就让该单元导电，否则就让该单元空着，重复这个过程，只要时间允许，以便估计对不同 p 的渗流概率。非渗流的图如图 P1.68(a)所示，渗流的图如图 P1.68(b)所示。本题更多的讨论请参阅 M. Schroeder 的 *Fractals*, *Chaos*, *Power Laws*(New York：W. H. Freeman,1991)。

1.69* 你是一个 TV 游戏秀的参与者，有三个相同的关闭的门导向三个房间，两个房间什么也没有，而第三个房间装有价值 100 000 美元的雷克萨斯豪华汽车，如果你选择对了，这辆豪华汽车就是你的。节目主持人(主持人是知道雷克萨斯豪华汽车在哪个房间)要求你选一个门。在你挑选一个门后，主持人打开一个门(不是你选的门)表明房间里没有雷克萨斯豪华汽车。证明：即使没有任何进一步的知识，将你原有选择的房间切换成剩下的那扇关闭的门的话，将大大地增加你赢得雷克萨斯豪华汽车的机会。

1.70 我们常常遇到在两个备选中确定一个最有可能的问题，在这种情况下，对于单个样本空间和事件域，我们给出两个概率测度，即$(\Omega, \mathcal{F}, P_1)$和$(\Omega, \mathcal{F}, P_2)$，要求确定在两种情况下观测到的事件 E 的概率，最有可能的选择是给出事件空间概率最大的那个。

考虑放在盒子里面的两枚硬币，一枚是均匀的，$P_1[\{H\}]=0.5$，另一枚是有偏的，$P_2[\{H\}]=p$，从盒子抽出一枚硬币，并将该硬币抛 10 次，我们只考虑把重复的抛硬币作为实验，所以样本空间为 $\Omega=$ {所有 H 和 T 的所有 10 个字符串}。我们观察事件 $E=$ {4 个 H 和 6 个 T 的总数}。

(a)观察事件 E 的两个概率 $P_1[E]$ 和 $P_2[E]$ 是多少?

(b)确定似然比 $L \triangleq P_1[E]/P_2[E]$ 与 p 的函数(当 $L>1$ 时，我们说均匀硬币最有可能，这样的检验称为似然比检验)。

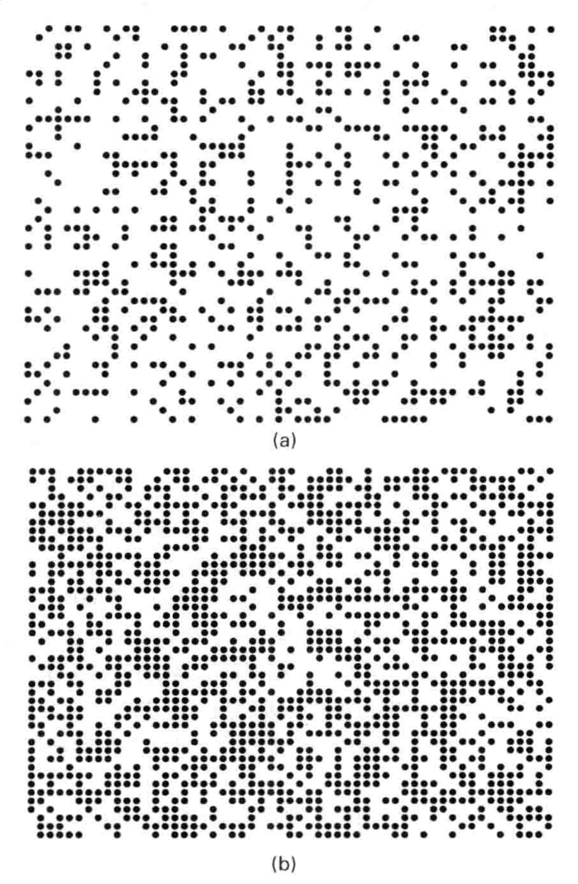


图 P1.68 (a)非渗流随机分形；(b)渗流随机分形。你能找出渗流路径吗?(摘自 *Fractals*, *Chaos*, *Power Laws*"(经M. Schroeder, W. H. Freeman, NewYork：,1991授权引用)

参考文献

1-1　E. Merzbacher, *Quantum Mechanics.* New York: John Wiley, 1961.

1-2　A. Kolmogorov, *Foundations of the Theory of Probability.* New York: Chelsea, 1950.

1-3　S. Pinker, *How the Mind Works.* New York: Norton, 1997.

1-4　B. O. Koopman, "The Axioms of Algebra and Intuitive Probability," *Annals of Mathematics* (2), Vol. 41, 1940, pp. 269-292.

1-5　A. Papoulis, and S. U. Pillai *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, 4th Ed, 2002.

1-6　R. Von Mises, *Wahrscheinlichkeit, Statistic and Wahrheit.* Vienna: Springer-Verlag, 1936.

1-7　W. B. Davenport, Jr., *Probability and Random Processes.* New York: McGraw-Hill, 1970.

1-8　W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd edition. New York: John Wiley, 1950, Chapter 2.

1-9　E. Parzen, *Modern Probability Theory and Its Applications.* New York: John Wiley, 1960, p. 119.

第2章 随机变量

2.1 引言

许多随机现象的实验结果是实数集：噪声电阻两端在时间 t 的电压 $v(t)$，电影院下一个观众的到达时间，在一个光脉冲里的光子数，电视屏幕上某一点的亮度，电灯泡失效前的开关次数，某个人的寿命，纽约到芝加哥火车上的人数，等等。所有这些情况的样本空间都是实轴上数的集合。

即使样本空间 Ω 不是数值的，我们也要产生一个新数值的样本空间 Ω，例如，将随机语音、颜色、灰度等转换成数字，将随机选择的一个人身体健康的状态转换成由体重、身高、血压、心率等组成的数值化“健康”向量，或者用一个向量描述受尽疾病折磨的患者的病情状况，如黑肺病患者，描述病情状况的向量由肺损伤的数量和大小以及肺部感染的数量两个分量组成。

在科学与工程中，无论考察的实验 $\mathscr{H}$ 是数值与否，几乎在所有情况下，都只对数值的实验结果感兴趣。为了得到数值的实验结果，我们需要一种从原始的样本空间 Ω 到实数直线 R^1 的规则或者映射，这种映射就是随机变量的基础，我们将在下面几节中详细进行讨论。

然而，应做一、二点注释和(a, b)随机变量的概念使我们用事件是一个数字的集合去代替原始的概率空间，这样的话，根据随机变量引入的概率空间，每一个事件都是 R^1 的一个子集。但是，R^1 的每一个子集都是一个事件吗？会不会存在一个 R^1 的子集给我们带来麻烦？如与概率的公理产生冲突？答案是“会”，但是，幸运的是，这些子集对工程和科学问题来说并不重要，我们说它们是不可测的①。在实际中重要的集合具有形式是$\{x=a\}$，$\{x: a\leqslant x\leqslant b\}$，$\{x: a<x\leqslant b\}$，$\{x: a\leqslant x<b\}$，$\{x: a<x<b\}$和它们的并集与交集，这 5 个区间很容易表示为$[a]$，$[a, b]$，$(a, b]$，$[a, b)$和$(a, b)$。包含两个端点的区间称为闭区间，不含两个端点的区间称为开区间。区间也有半闭(半开)的，如$(a, b]$是左开区间，或者称为右闭区间。由区间产生的 R^1 的子集的域在 1.4 节称为博雷尔域。

在同一个样本空间 Ω，我们可以定义多个随机变量，例如，假定 Ω 是由美国有代表性的一组人组成，令实验由随机选择一个人组成，X 表示人的寿命，Y 表示烟龄，人们可能要问，X 与 Y 有关系吗？也就是说我们能从 Y 预测 X 吗？假定定义第三个随机变量 Z，它表示人的体重，Z 与 X 或者 Y 有关系吗？

引入随机变量的主要优点在于我们能够定义概率函数，用概率函数可以容易计算不同事件的概率，这些函数自然必须与概率的公理一致。因此，在实数直线上定义事件时必须小心一点，在本节引入的思想将在下一节进行详细描述。

2.2 随机变量的定义

在样本空间 Ω 考虑一个实验 $\mathscr{H}$，Ω 中的点或者称为元素 ζ 是随机实验 $\mathscr{H}$ 的结果，如果对于每一个 ζ，我们赋予一个数 $X(\zeta)$，那么就在 ζ 与 R^1 实数直线之间建立了一种对应规则，在

---

① 参见附录 D 有关测度的简单讨论。

某些约束下的这种规则称为随机变量，简写为RV。这样，随机变量$X(\cdot)$或简写为X不是一个实际的变量，而是一个定义域为Ω的函数，它的值域是实数直线的一个子集。作为一个函数，X对每一个ζ产生一个特定的$X(\zeta)$，尽管对某个特定的$X(\zeta)$有可能是由多个实验结果ζ产生。现在考虑一个事件$E_B \subset \Omega(E_B \in \mathscr{F})$。

通过映射X，这一事件映射成实数直线上的点(参见图2.2-1)。特别是，事件$\{\zeta: X(\zeta) \leqslant x\}$常常缩写为$\{X \leqslant x\}$，以表示事件的重要性，我们应该给它赋予一个概率。作为实变量x的函数，概率$P[X \leqslant x] \triangleq F_X(x)$称为$X$的累积分布函数(Cumulative Distribution Function, CDF)，在参考文献[2-1]和参考文献[2-2]中证明了，要想$F_X(x)$与概率的公理化定义一致，函数X必须满足下列条件：对于每一个数的博雷尔集B，集合$\{\zeta: X(\zeta) \in B\}$必须与事件$E_B \in \mathscr{F}$对应，即它必须是概率测度$P$的定义域，说起来有点数学化，这要求在$R^1$中每一个博雷尔子集的$X$的逆像(组成了域$\mathscr{B}$①)都是一个事件，$X$才可能是一个随机变量。什么是逆像呢？考虑任意的博雷尔集B，$X(\zeta)$假定在B中，那么，在映射X时，Ω中点集E_B称为集合B的逆像。最终，在工程中感兴趣的所有集合都可以写成是形式为$(-\infty, x]$事件的可数的并集或交集。事件$\{\zeta: X(\zeta) \leqslant x\} \in \mathscr{F}$表示将$X$映射成$(-\infty, x] \in \mathscr{B}$。因此，如果$X$是一个随机变量，点集$(-\infty, x]$就是一个事件。

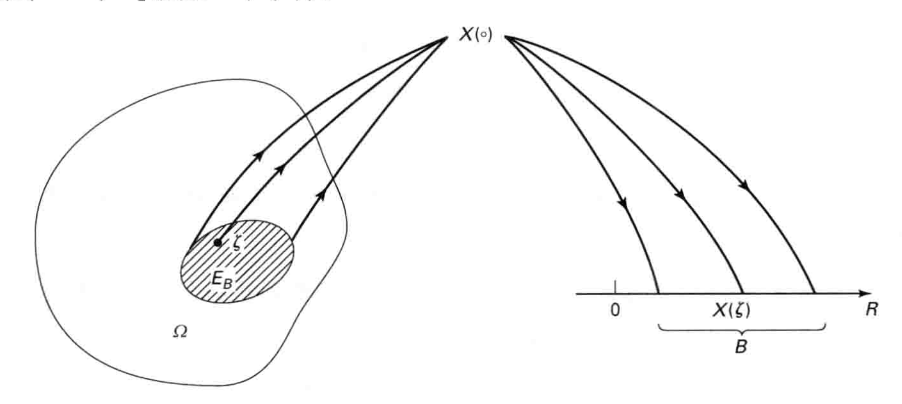


图2.2-1 随机变量定义的符号表示

在许多科学与工程的应用中，我们对X的实际形式或者集Ω的特性并不感兴趣，例如，可以设想实验为加热电阻并且观察电阻中电子的位置和速度，那么，集合Ω就是电阻中N个电子出现的所有位置和速度。令X表示由电阻产生的热噪声电流，很显然，$X: \Omega \to R^1$，尽管X的形式并没有指定，它是由一个相当复杂的量子电动力学方程，这一方程描述了由电子位置与速度到电流的映射关系。我们真正感兴趣的是X的行为。因此，尽管实验中隐含了样本空间Ω，我们的兴趣和在计算时用到的只是实数直线R^1和它的子集。在映射X时，实际的效果是产生了一个新的概率空间$\{R^1, \mathscr{B}, P_X\}$，其中$R^1$是实数直线，$\mathscr{B}$是博雷尔$\sigma$域，它是由半无穷区间$(-\infty, x]$的所有并集、交集和补集产生的，$P_X$是对每一个集合$A \in \mathscr{B}$赋予一个数的一簇函数②。

为了在$x = \pm\infty$处赋予函数$F_X(x)$期望的连续特性，我们要求事件$\{X = \infty\}$和$\{X = -\infty\}$有概率零，后者刻画了随机变量的完备性。以上讨论总结在下面的定义中。

① 定义在Ω上的事件的σ域用$\mathscr{F}$表示。在R^1上点一簇博雷尔子集用$\mathscr{B}$表示。它们的定义参见第1章的1.4节。

② 处理随机变量的一个很大的优势就是单个分段函数，即累积分布函数$F_X(x)$，能够取代集函数$P_X[\cdot]$，$P_X[\cdot]$要对每一个事件(集合)$A \in \mathscr{B}$进行描述是相当麻烦的，参见2.3节。

定义 2.2-1　令 $\mathscr{H}$ 为一随机实验，它的样本空间为 Ω，那么，实随机变量是一个定义在 Ω 上的函数，它满足下列几项：(i) 对于数 B 的博雷尔集，集合 $E_B \triangleq \{\zeta \in \Omega, X(\zeta) \in B\}$ 是一个事件；(ii) $P[X=-\infty]=P[X=+\infty]=0$。

一般认为，当 X 的范围是由可数点集组成，那么，X 就称为离散型随机变量，如果 X 的范围是连续的，X 就称为连续型随机变量。作为离散型和连续型随机变量的定义，这有点不太合适，因为 X 的取值范围常常要在整个实数直线 R^1 上，在 R^1 中那些以非零概率变换得不到的点是不可能事件①。

例 2.2-1　(随机人)从大街上随机选择一个人，询问他(她)是否有弟弟，如果回答是"no"，那么数据由随机变量 X 编码为 0，如果回答是"yes"，那么数据被编码为 1。那么，随机实验的样本空间为 $\Omega=\{\text{no}, \text{yes}\}$，$\sigma$ 域为 $\mathscr{F}=[\phi, \Omega, \{\text{no}\}, \{\text{yes}\}]$，概率 $P[\phi]=0$，$P[\Omega]=1$，$P[\text{no}]=3/4$(假设)，$P[\text{yes}]=1/4$。与 X 有关的概率是 $P[\phi]=0$，$P[X\leqslant\infty]=P[\Omega]=1$，$P[X=0]=P[\text{no}]=\dfrac{3}{4}$，$P[X=1]=P[\text{yes}]=\dfrac{1}{4}$，取任意的 x_1、x_2，例如，考虑 X 落在 $[x_1, x_2]$，$[x_1, x_2)$ 或 $(x_1, x_2]$ 的概率，于是

$$P[3 \leqslant X \leqslant 4] = P[\phi] = 0$$

$$P[0 \leqslant X < 1] = P[\text{no}] = \tfrac{3}{4}$$

$$P[0 \leqslant X \leqslant 2] = P[\Omega] = 1$$

$$P[0 < X \leqslant 1] = P[\text{yes}] = \tfrac{1}{4}$$

等等，每一个集合 $\{X=x\}$，$\{x_1\leqslant X\leqslant x_2\}$，$\{X\leqslant x_2\}$ 等都与定义在 Ω 上的事件有关，因此，X 是一个随机变量。

例 2.2-2　(公交车随机到达时间)公交车在 $[0, T]$ 随机到达，令 t 表示到达时间，那么，样本空间 Ω 为 $\Omega=\{t:t\in[0, T]\}$ 定义随机变量 X 为

$$X(t) = \begin{cases} 1, & t \in \left[\dfrac{T}{4}, \dfrac{T}{2}\right] \\ 0, & \text{其他} \end{cases}$$

假定到达时间在 $[0, T]$ 上是均匀的，计算 $P[X(t)=1]$ 或者 $P[X(t)=0]$ 或者 $P[X(t)\leqslant 5]$。

例 2.2-3　(从罐中抽取)一个罐装有 3 个彩色球，球分别为白色(W)、黑色(B)和红色(R)。实验为从罐中随机抽取一个球，那么，样本空间为 $\Omega=\{W, B, R\}$。随机变量 X 定义为

$$X(\zeta) = \begin{cases} \pi, & \zeta = W \text{ 或 } B \\ 0, & \zeta = R \end{cases}$$

计算概率 $P\{X\leqslant x_1\}$，其中 x_1 是任意数。于是，$\{X\leqslant 0\}=\{R\}$，$\{2\leqslant X<4\}=\{W, B\}$ 有关概率的计算留做一个练习。

例 2.2-4　(机会轮)一个转轮指向 50 个扇区，每个扇区分别标有 $n=0, 1, \cdots, 49$，实验为转动轮子。由于游戏者只对偶数或奇数结果感兴趣，那么，选择 $\Omega=\{\text{偶数}, \text{奇数}\}$，在 σ 域中只有事件 $\{\phi, \Omega, \text{偶数}, \text{奇数}\}$。令 $X=n$，即 n 是指针指向的那个数字，则 X 就取那个值，X

① 另一个定义如下：如果 $F_X(x)$ 是一个阶梯函数，那么 X 就是离散型随机变量，如果 $F_X(x)$ 是一个连续函数，那么 X 就是连续型随机变量。有些随机变量不能分类为离散型或连续型随机变量，它们将在 2.5 节中讨论。

是随机变量吗？注意，集合{2, 3}的逆像不是一个事件，因此，X 不是这个概率空间上的一个有效随机变量，因为它不是 Ω 的函数。

2.3 累积分布函数

在例2.2-1中导出了 X 的事件空间包括{0, 1}，{0}，{1}，ϕ，其概率为 $P\{X=0$ 或 $1]=1$，$P[X=0]=3/4$，$P[X=1]=1/4$，$P[\phi]=0$，由这些概率，我们能够导出任何其他概率，如 $P[X\leqslant 0.5]$。在许多情况下，为每个事件写出概率 $P[\cdot]$ 是非常棘手的。由于这一原因，我们引入概率函数，称为累积分布函数。累积分布函数是 x 的函数，它包含了计算博雷尔域中任何事件 E 的概率 $P[E]$ 的所有必须的信息。CDF，即 $F_X(x)$，定义为

$$F_X(x)=P[\{\zeta\colon X(\zeta)\leqslant x\}]=P_X[(-\infty,x]] \tag{2.3-1}$$

式(2.3-1)读为"样本空间中函数值 $X(\zeta)$ 小于等于 x 的所有实验结果 ζ 的集合"，因此，存在一个实验结果的子集 $\{\zeta\colon X(\zeta)\leqslant x\}\subset\Omega$，映射 $X(\cdot)$ 产生一个集合 $[-\infty, x]\subset R^1$ 集合 $\{\zeta\colon X(\zeta)\leqslant x\}\subset\Omega$ 和 $[-\infty, x]\subset R^1$ 是一个等价事件。通常忽略与样本空间的关系，只写为 $P[X\leqslant x]$ 或 $P[a<X\leqslant b]$。

通常用大写字母表示随机变量，如 X，Y，Z，用小写字母表示随机变量的值，如 x，y，z。$F_X(x)$ 中的下标 X 与 CDF 描述的随机变量有关，因此，$F_X(y)$ 表示随机变量 X 的 CDF 在实数 y 处的值，它等于概率 $P[X\leqslant y]$。如果 $F_X(x)$ 在某点不连续(x_o 点)，那么 $F_X(x_o)$ 将取 CDF 在 x_o 的右边的值(称为右连续)。

$F_X(x)$ 的性质①

(i) $F_X(\infty)=1$，$F_X(-\infty)=0$。

(ii) $x_1\leqslant x_2\rightarrow F_X(x_1)\leqslant F_X(x_2)$，即 $F_X(x)$ 是 x 的不减函数。

(iii) $F_X(x)$ 是右连续的，即

$$F_X(x)=\lim_{\varepsilon\to 0}F_X(x+\varepsilon),\qquad \varepsilon>0$$

(ii) 的证明 考虑事件 $\{x_1<X\leqslant x_2\}$，其中 $x_2>x_1$，集合 $[x_1, x_2]$ 是非空集，且 $\in\mathscr{B}$，因此

$$0\leqslant P[x_1<X\leqslant x_2]\leqslant 1$$

而

$$\{X\leqslant x_2\}=\{X\leqslant x_1\}\cup\{x_1<X\leqslant x_2\}$$

和

$$\{X\leqslant x_1\}\cap\{x_1<X\leqslant x_2\}=\phi$$

因此

$$F_X(x_2)=F_X(x_1)+P[x_1<X\leqslant x_2]$$

即

$$P[x_1<X\leqslant x_2]=F_X(x_2)-F_X(x_1)\geqslant 0,\qquad x_2>x_1 \tag{2.3-2}$$

---

① 性质(i)和性质(iii)需要证明，这要借助第8章中扩展的公理来完成，也可参阅 Davenport[2-3, Chapter 4]。

下列结果留给读者自己去证明：

$$P[a \leqslant X \leqslant b] = F_X(b) - F_X(a) + P[X = a]$$
$$P[a < X < b] = F_X(b) - P[X = b] - F_X(a)$$
$$P[a \leqslant X < b] = F_X(b) - P[X = b] - F_X(a) + P[x = a]$$

例 2.3-1　(校验位)实验为观察计算机存储器中的一个字的校验位的电压 X，如果位是 on，那么 $X=1$，如果位是 off，那么 $X=0$。假定 off 状态的概率为 q，on 状态的概率为 $1-q$，那么，样本空间只有两个点：$\Omega=\{\text{off}, \text{on}\}$。

$F_X(x)$的计算

(i) $x<0$，事件$\{X\leqslant x\}=\phi$，所以，$F_X(x)=0$。

(ii) $0\leqslant x<1$：事件$\{X\leqslant x\}$等价于事件$\{\text{off}\}$，排斥事件$\{\text{on}\}$。

$$X(\text{on}) = 1 > x$$
$$X(\text{off}) = 0 \leqslant x$$

因此，$F_X(x)=q$。

(iii) $x\geqslant 1$ ：事件$\{X\leqslant x\}$是必然事件，因为

$$X(\text{on}) = 1 \leqslant x$$
$$X(\text{off}) = 0 \leqslant x$$

结果如图 2.3-1 所示。

例 2.3-2　(等公交车)公交车在$(0, T]$时间随机到达，令随机变量 X 表示到达时间，很显然，当 $t\leqslant 0$ 时，$F_X(t)=0$，而 $F_X(T)=1$，因为前者是一个不可能事件的概率，后者是必然事件的概率。假定公交车是等均匀可能地在$(0, T]$间隔的任何时间到达，那么

$$F_X(t) = \begin{cases} 0, & t \leqslant 0 \\ \dfrac{t}{T}, & 0 < t \leqslant T \\ 1, & t > T \end{cases} \tag{2.3-3}$$

实际上，式(2.3-3)定义了“等可能”，CDF 如图 2.3-2 所示，在这种情况下，我们说 X 是服从均匀分布的。

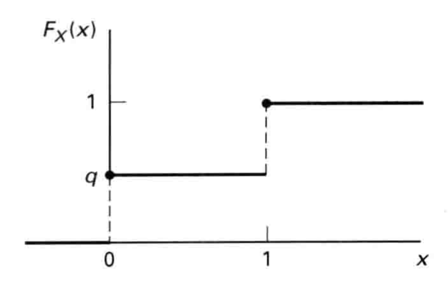


图 2.3-1　校验位随机实验的累积分布函数

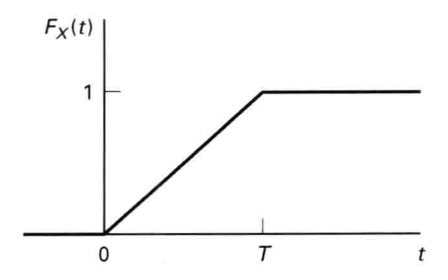


图 2.3-2　例 2.3-2 中均匀分布随机变量 X 的累积分布函数

如果 $F_X(x)$是 x 的连续函数，那么

$$F_X(x) = F_X(x^-) \tag{2.3-4}$$

但是，如果 $F_X(x)$在 x 点是不连续的，那么，由式(2.3-2)

$$\begin{aligned} F_X(x) - F_X(x^-) &= P[x^- < X \leqslant x] \\ &= \lim_{\varepsilon \to 0} P[x - \varepsilon < X \leqslant x] \\ &\triangleq P[X = x] \end{aligned} \tag{2.3-5}$$

通常$P[X=x]$是x的不连续函数，当$F_X(x)$是连续的时，$P[X=x]$为零，只有在$F_X(x)$不连续处才为非零。

例2.3-3 （二项式分布函数）计算参数为(n, p)的二项式随机变量X的CDF。

解 由于X只取离散值，即$X \in \{0, 1, 2, \cdots, n\}$，事件$\{X \leqslant x\}$与$\{X \leqslant [x]\}$是相同的，其中$[x]$是等于或小于$x$的最大整数。那么，$F_X(x)$由阶梯函数表示为

$$F_X(x) = \sum_{j=0}^{[x]} \binom{n}{k} p^j (1-p)^{n-j}$$

当$p=0.6$，$n=4$时，CDF是一个楼梯形函数，如图2.3-3所示。

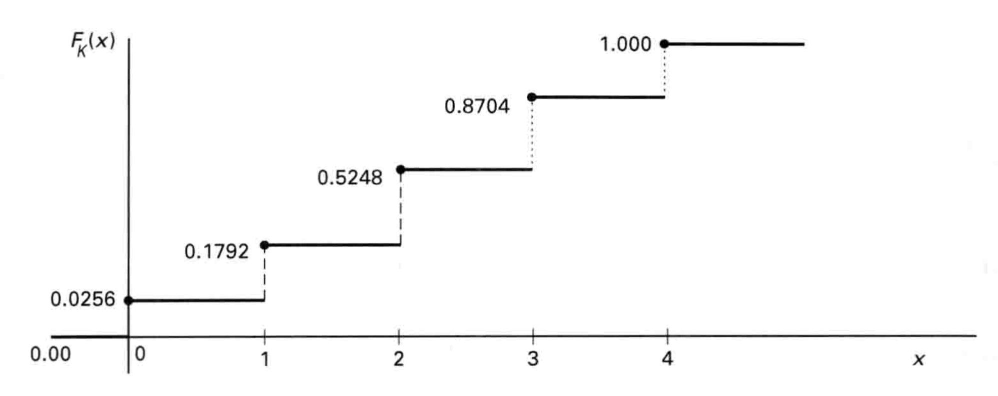


图2.3-3 $p=0.6$、$n=4$的二项式随机变量的累积分布函数

例2.3-4 （计算二项式概率）利用例2.3-3计算下列概率。

(a) $P[1.5 < X < 3]$

(b) $P[0 \leqslant X \leqslant 3]$

(c) $P[1.2 < X \leqslant 1.8]$

(d) $P[1.99 \leqslant X < 3]$

解

(a) $P[1.5 < X < 3] = F_X(3) - P[X=3] - F_X(1.5) = 0.8704 - 0.3456 - 0.1792 = 0.3456$

(b) $P[0 \leqslant X \leqslant 3] = F_X(3) - F_X(0) + P[X=0] = 0.8704 - 0.0256 + 0.0256 = 0.8704$

(c) $P[1.2 < X \leqslant 1.8] = F_X(1.8) - F_X(1.2) = 0.1792 - 0.1792 = 0$

(d) $P[1.99 \leqslant X < 3] = F_X(3) - P[X=3] - F_X(1.99) + P[X=1.99] = 0.8704 - 0.3456 - 0.1792 + 0 = 0.3456$

注意到即使是离散型随机变量，CDF也有可能取连续变量的函数。然而，对于一个离散型随机变量，有时简单地认为CDF也是离散的。令X是一个离散随机变量，取值为$\{x_k\}$，概率质量函数(PMF)为$P_X(x_k)$。那么，离散CDF也只定义在$\{x_k\}$。假定这些值是递增的，即对于所有的k，$x_k < x_{k+1}$，离散CDF为

$$F_X(x_k) \triangleq \sum_{j=-\infty}^{k} P(x_j) \text{ 对所有 } k$$

在这个式子里，我们只是在对应于样本空间可数的实验结果上计算 CDF。

再看上面的计算二项式 $b(n, p)$ 的例子，但采用离散 CDF，我们说随机变量 k 在集合 $\{0 \leqslant k \leqslant 4\}$ 上取值，它的离散 CDF 为

$$F_K(k) = \sum_{j=0}^{k} (0.6)^j (1-0.6)^{n-j}, \qquad 0 \leqslant k \leqslant 4$$

读者可能注意到，离散 CDF 不能用来计算像 $P[1.5 < K < 3]$ 这样的概率，因为它不能在 1.5 处进行计算。由于这样的原因，即使随机变量是离散的，也通常认为 CDF 定义在连续域。

2.4 概率密度函数(pdf)

如果 $F_X(x)$ 是连续的、可微的，那么，pdf 定义为

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} \tag{2.4-1}$$

性质 如果 $f_X(x)$ 存在，那么

(i) $$f_X(x) \geqslant 0 \tag{2.4-2}$$

(ii) $$\int_{-\infty}^{\infty} f_X(\zeta)\mathrm{d}\zeta = F_X(\infty) - F_X(-\infty) = 1 \tag{2.4-3}$$

(iii) $$F_X(x) = \int_{-\infty}^{x} f_X(\zeta)\mathrm{d}\zeta = P[X \leqslant x] \tag{2.4-4}$$

(iv) $$F_X(x_2) - F_X(x_1) = \int_{-\infty}^{x_2} f_X(\zeta)\mathrm{d}\zeta - \int_{-\infty}^{x_1} f_X(\zeta)\mathrm{d}\zeta = \int_{x_1}^{x_2} f_X(\zeta)\mathrm{d}\zeta = P[x_1 < X \leqslant x_2] \tag{2.4-5}$$

$f_X(x)$ 的解释

$$P[x < X \leqslant x + \Delta x] = F_X(x + \Delta x) - F_X(x)$$

如果 $F_X(x)$ 的一阶导数是连续的，那么，对于一个充分小的 Δx

$$F_X(x + \Delta x) - F_X(x) = \int_{x}^{x+\Delta x} f(\zeta)\mathrm{d}\zeta \approx f_X(x)\Delta x$$

因此，对于小的 Δx

$$P[x < X \leqslant x + \Delta x] \approx f_X(x)\Delta x \tag{2.4-6}$$

注意到，如果 $f_X(x)$ 存在，也意味着它是有界的，且至多有有限个不连续点，那么 $F_X(x)$ 是连续的。因此，由式(2.3-5)，$P[X = x\} = 0$。

单变量正态(高斯[①])pdf pdf 为

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}, -\infty < x < +\infty \tag{2.4-7}$$

① 德国数学家/物理学家 Carl F. Gauss (1777 - 1855)。

其中有两个不同的参数：均值μ和标准差σ（大于0）（注意：σ^2称为方差）。对所有x积分就可以证明这个密度是有效的。

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\mathrm{d}x$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{y^2}{2}}\mathrm{d}y,\quad \text{通过变量替换 } y\triangleq\frac{x-\mu}{\sigma}$$
$$=\frac{2}{\sqrt{2\pi}}\int_{0}^{+\infty}\mathrm{e}^{-\frac{y^2}{2}}\mathrm{d}y=\frac{2}{\sqrt{2\pi}}\sqrt{\frac{\pi}{2}}=\frac{2\sqrt{\pi}}{2\sqrt{\pi}}=1$$

其中我们利用了已知积分

$$\int_0^{\infty}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x=\sqrt{\frac{\pi}{2}}$$

目前高斯（正态）随机变量在应用中是非常常见的，用一个特殊的表示形式来表示它，我们说X服从$N(\mu,\sigma^2)$或者写成X：$N(\mu,\sigma^2)$来表示这一分布。

对于任意的一个随机变量，通常可以用下面两式计算均值和方差（标准差的平方）

$$\mu\triangleq\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x \tag{2.4-8}$$

和

$$\sigma^2\triangleq\int_{-\infty}^{\infty}(x-\mu)^2f_X(x)\mathrm{d}x \tag{2.4-9}$$

将在第4章证明高斯分布中的这两个参数就是高斯随机变量真正的均值和方差。

对于一个离散随机变量，可以从求和计算均值和方差

$$\mu\triangleq\sum_{i=-\infty}^{\infty}x_iP_X(x_i) \tag{2.4-10}$$

和

$$\sigma^2\triangleq\sum_{i=-\infty}^{\infty}(x_i-\mu)^2P_X(x_i) \tag{2.4-11}$$

下面给出均值和方差的一些简单计算例子。

例2.4-1 令$f_X(x)=1$，其中$0<x\leq1$，其他地方为零。这个pdf是下面要讨论的均匀分布的一种特殊情况。均值为

$$\mu=\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x=\int_0^1 x\,\mathrm{d}x=0.5$$

方差为

$$\sigma^2=\int_{-\infty}^{\infty}(x-\mu)^2f_X\mathrm{d}x=\int_0^1(x-0.5)^2\mathrm{d}x=1/12$$

例2.4-2 假定$P_X(0)=P_X(2)=0.25$，$P_X(1)=0.5$，其他值为零，对于这一离散随机变量，利用式(2.4-10)和式(2.4-11)，得

$$\mu=0\times0.25+1\times0.5+2\times0.25=1$$

和

$$\sigma^2 = (0-1)^2 \times 0.25 + (1-1)^2 \times 0.5 + (2-1)^2 \times 0.25 = 0.5$$

均值和方差是统计矩的常见例子，统计矩将放到第 4 章讨论，正态 pdf 如图 2.4-1 所示。

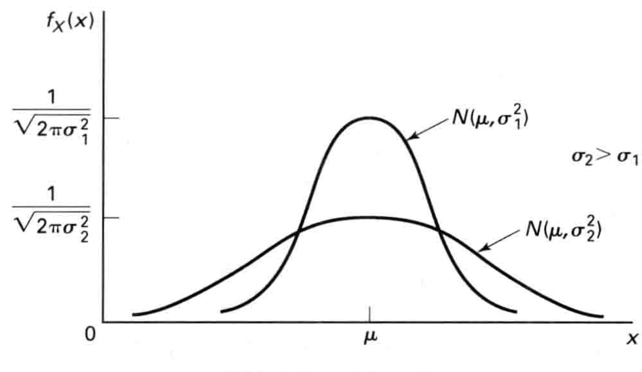


图 2.4-1 正态 pdf

正态 pdf 在科学与工程的许多分支领域以及社会和人口统计学研究中有广泛应用。例如，儿童的 IQ、男人(或女人)的身高、由热引起的电阻噪声电压等都可以在一个大的取值范围内用正态近似。

高斯 pdf 到标准正态分布的转换 假定 X: $N(\mu, \sigma^2)$，计算 $P[a < X \leqslant b]$。我们有

$$P[a < X \leqslant b] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_a^b \mathrm{e}^{-\frac{1}{2}\left[\frac{x-u}{\sigma}\right]^2} \mathrm{d}x$$

用 $\beta \triangleq (x-\mu)/\sigma$，$\mathrm{d}\beta = (1/\sigma)\mathrm{d}x$，$b' \triangleq (b-u)/\sigma$，$a' \triangleq (a-\mu)/\sigma$，得

$$P[a < X \leqslant b] = \frac{1}{\sqrt{2\pi}} \int_{a'}^{b'} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{b'} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x - \frac{1}{\sqrt{2\pi}} \int_0^{a'} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x$$

函数

$$\mathrm{erf}(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_0^x \mathrm{e}^{-\frac{1}{2}t^2} \mathrm{d}t \qquad (2.4\text{-}12)$$

有时也称为误差函数[erf(x)]，尽管误差函数 erf(x) 也有其他形式的定义①。误差函数 erf(x) 的取值列在表 2.4-1 中，图 2.4-2 给出了误差函数的图形。

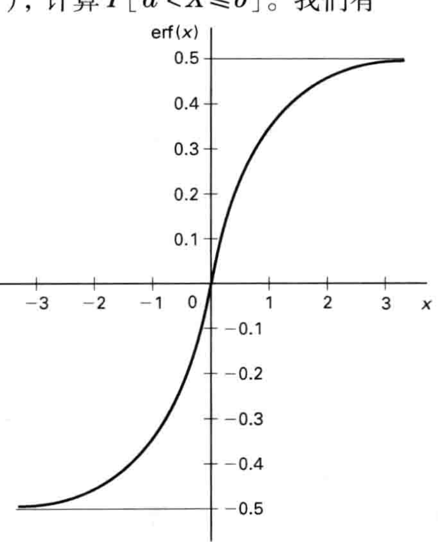


图 2.4-2 erf(x) 与 x 的曲线

因此，如果 X:$N(\mu, \sigma^2)$，那么

$$P[a < X \leqslant b] = \mathrm{erf}\left(\frac{b-\mu}{\sigma}\right) - \mathrm{erf}\left(\frac{a-\mu}{\sigma}\right) \qquad (2.4\text{-}13)$$

① 例如，误差函数用得最广泛的定义是 $\mathrm{erf}_2(x) \triangleq (2/\sqrt{\pi})\int_0^x \mathrm{e}^{-t^2}\mathrm{d}t$，MATLAB 采用这一定义，两个误差函数的关系是 $\mathrm{erf}(x) = \frac{1}{2}\mathrm{erf}_2(x/\sqrt{2})$。

表2.4-1 erf(x)的值

$$\mathrm{erf}(x)=\frac{1}{2\pi}\int_0^x \exp\left(-\frac{1}{2}t^2\right)\mathrm{d}t$$

x	erf(x)	x	erf(x)	x	erf(x)	x	erf(x)
0.05	0.01994	1.05	0.35314	2.05	0.47981	3.05	0.49884
0.10	0.03983	1.10	0.36433	2.10	0.48213	3.10	0.49902
0.15	0.05962	1.15	0.37492	2.15	0.48421	3.15	0.49917
0.20	0.07926	1.20	0.38492	2.20	0.48609	3.20	0.49930
0.25	0.09871	1.25	0.39434	2.25	0.48777	3.25	0.49941
0.30	0.11791	1.30	0.40319	2.30	0.48927	3.30	0.49951
0.35	0.13683	1.35	0.41149	2.35	0.49060	3.35	0.49958
0.40	0.15542	1.40	0.41924	2.40	0.49179	3.40	0.49965
0.45	0.17364	1.45	0.42646	2.45	0.49285	3.45	0.49971
0.50	0.19146	1.50	0.43319	2.50	0.49378	3.50	0.49976
0.55	0.20884	1.55	0.43942	2.55	0.49460	3.55	0.49980
0.60	0.22575	1.60	0.44519	2.60	0.49533	3.60	0.49983
0.65	0.24215	1.65	0.45052	2.65	0.49596	3.65	0.49986
0.70	0.25803	1.70	0.45543	2.70	0.49652	3.70	0.49988
0.75	0.27337	1.75	0.45993	2.75	0.49701	3.75	0.49990
0.80	0.28814	1.80	0.46406	2.80	0.49743	3.80	0.49992
0.85	0.30233	1.85	0.46783	2.85	0.49780	3.85	0.49993
0.90	0.31594	1.90	0.47127	2.90	0.49812	3.90	0.49994
0.95	0.32894	1.95	0.47440	2.95	0.49840	3.95	0.49995
1.00	0.34134	2.00	0.47724	3.00	0.49864	4.00	0.49996

例2.4-3 （电阻的公差）假定我们从一批 $\mu=1000\ \Omega$，$\sigma=200\ \Omega$ 的电阻中选择一个电阻 R，R 的值在 900～1100 Ω 之间的概率是多少？

解： 假定 R:$N[1000,(200)^2]$，由式(2.4-13)得

$$P[900<R\leqslant 1100]=\mathrm{erf}(0.5)-\mathrm{erf}(-0.5)$$

而 $\mathrm{erf}(-x)=-\mathrm{erf}(-x)$[由式(2.4-12)得出]，因此

$$P[900<R\leqslant 1100]=0.38$$

利用图2.4-3可以辅助进行推导，对于 X: $N(0,1)$，很容易导出下面的式子。假定 $x>0$，那么

$$P[X\leqslant x]=\frac{1}{2}+\mathrm{erf}(x) \tag{2.4-14a}$$

$$P[X>-x]=\frac{1}{2}+\mathrm{erf}(x) \tag{2.4-14b}$$

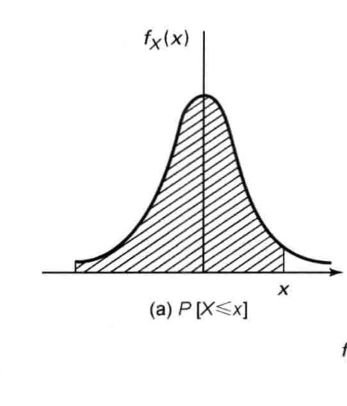
$$P[X>x]=\frac{1}{2}-\mathrm{erf}(x) \tag{2.4-14c}$$

$$P[-x<X\leqslant x]=2\,\mathrm{erf}(x) \tag{2.4-14d}$$

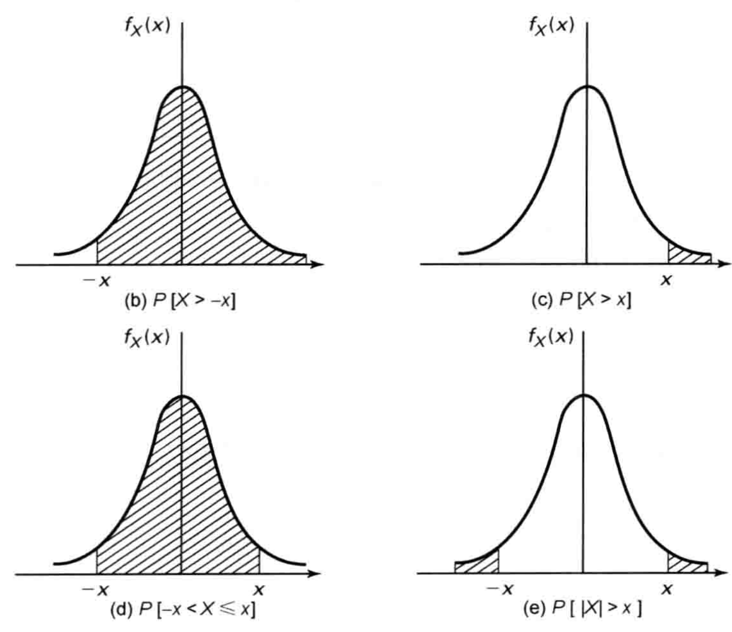
$$P[|X|>x]=1-2\,\mathrm{erf}(x) \tag{2.4-14e}$$

例2.4-4 (制造)一个金属拉杆应为1 m长，但由于制造的不完善，实际长度L是一个均值$\mu=1$、标准差为$\sigma=0.005$的高斯随机变量，拉杆长度位于[0.99, 1.01]的概率是多少？由于随机变量L: $N(1,(0.005)^2)$，我们有

$$
\begin{aligned}
P[0.99 < L \leqslant 1.01] &= \int_{0.99}^{1.01} \frac{1}{\sqrt{2\pi}(0.005)} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-1.00}{0.005}\right)^2} \mathrm{d}x \\
&= \int_{\frac{0.99-1.00}{0.005}}^{\frac{1.01-1.00}{0.005}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x \\
&= \int_{-2}^{2} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x \\
&= 2\,\mathrm{erf}(2) \doteq 2 \times 0.4772 \qquad (\text{参见表 2.4-1}) \\
&= 0.954
\end{aligned}
$$



(a) $P[X \leqslant x]$



(b) $P[X > -x]$　(c) $P[X > x]$　(d) $P[-x < X \leqslant x]$　(e) $P[|X| > x]$

图2.4-3　曲线下阴影部分的区域

其他四种常用密度函数

1. 瑞利分布(Rayleigh)($\sigma>0$)

$$f_X(x)=\frac{x}{\sigma^2}\mathrm{e}^{-x^2/2\sigma^2}u(x) \tag{2.4-15}$$

其中，连续单位阶跃函数定义为

$$u(x)\triangleq\begin{cases}1, & 0\leqslant x<\infty\\0, & -\infty<x<0\end{cases}$$

于是，对于 $x<0$，$f_X(x)=0$。火箭着陆误差、某些波形幅度的随机起伏、步枪射击时子弹脱离靶心的径向分布都服从瑞利分布。

2. 指数分布($\mu>0$)

$$f_X(x)=\frac{1}{\mu}\mathrm{e}^{-x/\mu}u(x) \tag{2.4-16}$$

等待时间、机器的寿命、不相干光的密度变化等都可以用指数分布来描述。

3. 均匀分布($b>a$)

$$\begin{aligned}f_X(x)&=\frac{1}{b-a}, \quad a<x<b\\&=0 \quad \text{其他}\end{aligned} \tag{2.4-17}$$

均匀分布应用于通信理论、排队模型等。此外，除端点外对实验结果的分布没有任何先验知识的时候，即我们并不知道商务电话什么时候来、但又肯定要来，比如在上午 9 点到下午 5 点之间，这时就可以用均匀分布来描述。有时也用 $U(a, b)$ 表示均匀分布，下边界为 a，上边界为 b。

三种 pdf 如图 2.4-4 所示。

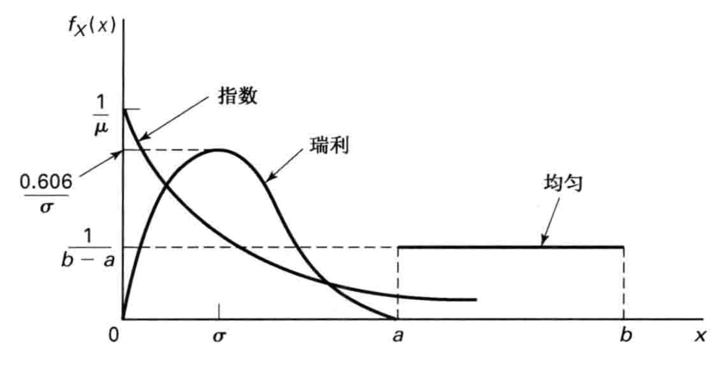


图 2.4-4　瑞利、指数和均匀 pdf

4. 拉普拉斯分布 pdf 定义如下：

$$f_X(x)=\frac{c}{2}\mathrm{e}^{-c|x|}, \ -\infty<x<\infty, \ c>0 \tag{2.4-18}$$

拉普拉斯分布广泛应用于语音和图像处理中，可用于对邻近样本差建模，或者用于对样本点和它的近邻信号电平差建模。由于样本点电平和它的近邻相同，所以拉普拉斯峰值在零点。拉普拉斯 pdf 有时也写成

$$f_X(x)=\frac{1}{\sqrt{2}\sigma}\exp[-\sqrt{2}|x|/\sigma], \ -\infty<x<\infty, \ \sigma>0 \tag{2.4-19}$$

其中 σ 是拉普拉斯随机变量 X 的标准差，在第 4 章再进行解释。拉普拉斯 pdf 如图 2.4-5 所

示。在图像压缩中，拉普拉斯模型对所谓的“AC 系数”是合适的，“AC 系数”出现在称为 DCT① 的去相关变换中之后。DCT 变换应用到 8 ×8 的像素块。

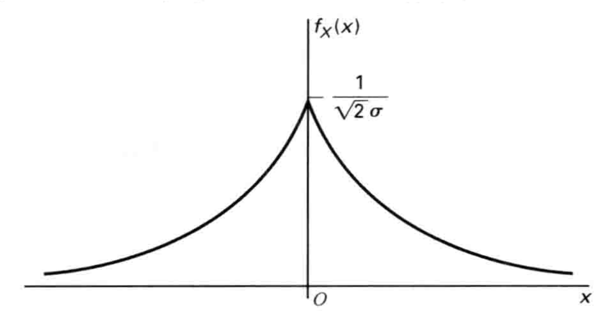


图 2.4-5　在语音和图像的计算机分析中使用拉普拉斯 pdf

例 2.4-5　(辐射功率)我们发现蜂窝电话在某个距离接收基站的接收功率服从参数为 $\sigma=1$ mW的瑞利分布。功率 W 小于 0.8 mW 的概率是多少？由于功率可以用瑞利随机变量建模，因此

$$
\begin{aligned}
P[W \leqslant 0.8] &= \int_{0.0}^{0.8} x\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x, \quad \text{由于 } \sigma^2 = 1 \\
&= \int_{0.0}^{0.8} \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}(\frac{1}{2}x^2) \\
&= \int_{0.0}^{0.32} \mathrm{e}^{-y}\mathrm{d}y, \quad \text{做变量替换 } y \triangleq \frac{1}{2}x^2 \\
&= 1 - \mathrm{e}^{-0.32} \approx 0.29
\end{aligned}
$$

例 2.4-6　(图像压缩)在为 JPEG 图像压缩系统设计量化器时，我们需要知道变换 AC 系数的范围。这个系数 X 用参数为 σ 的拉普拉斯模型来描述，事件 $\{|X|>k\sigma\}$ 的概率是多少？这个概率表示为 k 的函数，其中 $k=1, 2, 3, \cdots$。如果选择一个 k 使这个概率足够低，那么在 $[-k\sigma, +k\sigma]$ 范围内设计量化器，就只有少量的情况出现饱和。我们需要计算

$$
\begin{aligned}
P[|X| > k\sigma] &= \int_{k\sigma}^{\infty} \frac{1}{\sqrt{2}\sigma} \exp\left(-\sqrt{2}x/\sigma\right)\mathrm{d}x + \int_{-\infty}^{-k\sigma} \frac{1}{\sqrt{2}\sigma} \exp\left(+\sqrt{2}x/\sigma\right)\mathrm{d}x \\
&= 2\int_{k\sigma}^{\infty} \frac{1}{\sqrt{2}\sigma} \exp\left(-\sqrt{2}x/\sigma\right)\mathrm{d}x \\
&= 2\int_{k}^{\infty} \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}y\right)\mathrm{d}y, \quad y \triangleq x/\sigma \\
&= 2\left[\frac{1}{2}\exp\left(-\sqrt{2}y\right)\right]\Bigg|_k^{\infty} \\
&= \exp\left(-\sqrt{2}k\right)
\end{aligned}
$$

当 $k=2$ 时，得到概率为 0.059，当 $k=5$ 时，概率为 0.85×10^{-3}，即在 1000 个系数中大约有一个出现饱和。

① DCT 表示离散余弦变换，它是一种用在信号分析中的 DFT 的一种变形，二维 DCT 用于图像处理，它是由一个对行的一维 DCT 跟着一个对列的一维 DCT 组成。

在表2.4-2列出了一些常见的连续型随机变量的概率密度和分布函数。

表2.4-2 常见的连续型随机变量的概率密度和分布函数

分布类型	pdf $f_X(x)$	CDF $F_X(x)$
均匀$U(a,b)$	$\frac{1}{b-a}[u(x-a)-u(x-b)]$	$\begin{cases}0, & x<a,\\ \frac{x-a}{b-a}, & a\leqslant x<b\\ 1, & b\leqslant x\end{cases}$,
指数$\mu>0$	$\frac{1}{\mu}\mathrm{e}^{-x/\mu}\,u(x)$	$\begin{cases}0, & x<0,\\ 1-\mathrm{e}^{-x/\mu}, & x\geqslant 0\end{cases}$
高斯$N(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2]$	$\frac{1}{2}+\mathrm{erf}(\frac{x-\mu}{\sigma})$
拉普拉斯$\sigma>0$	$\frac{1}{\sqrt{2}\sigma}\exp[-\sqrt{2}\lvert x\rvert/\sigma]$	$\frac{1}{2}[1+\mathrm{sgn}(x)(1-\exp(-\sqrt{2}\lvert x\rvert/\sigma))]$
瑞利$\sigma>0$	$\frac{x}{\sigma^2}\mathrm{e}^{-x^2/2\sigma^2}u(x)$	$\left[1-\mathrm{e}^{-x^2/2\sigma^2}\right]u(x)$

更多的高级分布

5. χ^2分布(n为整数$n\geqslant1$)

$$f_X(x)=K_\chi x^{\left(\frac{n}{2}\right)-1}\mathrm{e}^{-\frac{x}{2}}u(x) \tag{2.4-20}$$

其中归一化常数K_χ为$K_\chi=\dfrac{1}{2^{n/2}\Gamma(n/2)}$，$\Gamma(\cdot)$为附录B中讨论的伽马函数。$\chi^2$ pdf如图2.4-6所示。

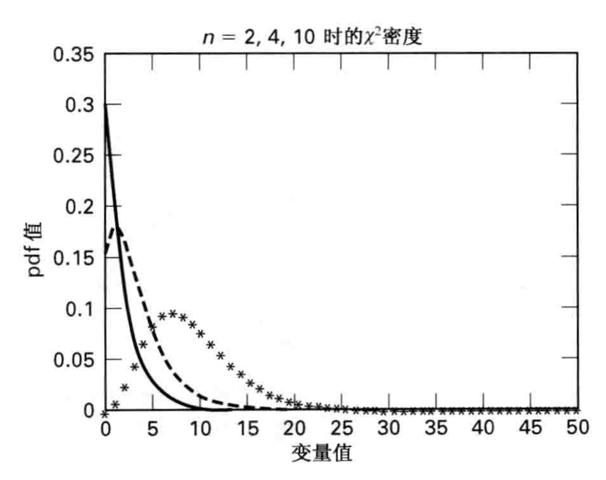


图2.4-6 $n=2$(实线)、$n=4$(虚线)和$n=10$(星号)时的χ^2概率密度。注意到n越大，概率密度的形状越趋于具有正的均值参数μ的正态概率密度

6. 伽马分布($b>0$, $c>0$)

$$f_X(x)=K_\gamma x^{b-1}\mathrm{e}^{-cx}u(x) \tag{2.4-21}$$

若伽马pdf的推广存在，则$K_\gamma=c^b/\Gamma(b)$。

7. 学生t分布(n为整数)

$$f_X(x)=K_{st}\left(1+\frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)},-\infty<x<\infty \tag{2.4-22}$$

其中

$$K_{st}=\frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{\pi n}}$$

χ^2 和学生 t 分布在统计学中被广泛应用①，我们在本书后面会遇到这些分布。伽马分布是其他分布的母分布，例如，$b=1$ 时对应的伽马分布为指数分布，$b=n/2$ 和 $c=1/2$ 时的伽马分布对应于 χ^2 分布。

8. 贝塔分布($\alpha\geqslant 0$, $\beta\geqslant 0$)

$$f_X(x;\alpha,\beta)=\begin{Bmatrix}\frac{(\alpha+\beta+1)!}{\alpha!\beta!}x^{\alpha}(1-x)^{\beta},\ 0<x<1\\ 0\text{，其他}\end{Bmatrix}$$

贝塔分布是统计学中的一种双参数的分布，如图 2.4-7 所示。

在工程和科学中还有其他一些重要的分布，我们将在今后的概率论的研究中遇到。然而，它们都有一些共同的性质。

$$f_X(x)\geqslant 0 \tag{2.4-23}$$

$$\int_{-\infty}^{\infty}f_X(x)\mathrm{d}x=1 \tag{2.4-24}$$

当 $F_X(x)$ 不连续时，严格地说它的某些点的导数并不存在，因此，pdf 也是不存在的。在描述 X 时，概率函数是否有用依赖于 X 的分类，接下来将讨论这个问题。

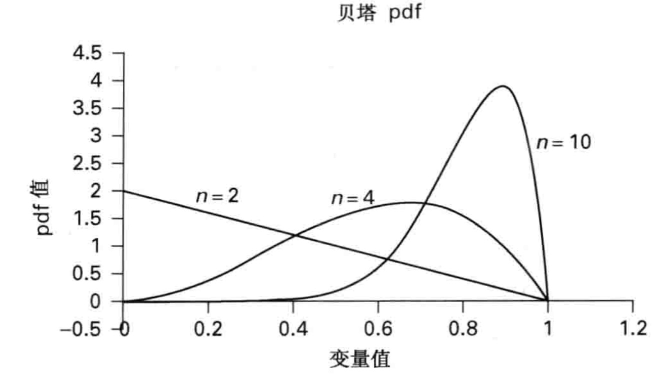


图 2.4-7 $\beta=1$、$\alpha=n-2$、对应于各种 n 值的贝塔 pdf。当 $\beta=0$、$\alpha=0$ 时，贝塔pdf变成了 $0<x<1$ 上的均匀分布

2.5 连续型、离散型和混合型随机变量

如果 $F_X(x)$ 对每个 x 都是连续的、且它的导数除某些可数点外是可导的，那么就称 X 是连续型随机变量，在 $F'_X(x)$ 存在的 x 点处，pdf 是 $f_X(x)F'_X(x)$。在 $F_X(x)$ 连续而 $F'_X(x)$ 不连续的点，我们可以给 $f_X(x)$ 赋任意的一个正数；那么 $f_X(x)$ 对每一个 x 都有定义，就可以使用下面的重要公式：

① 之所以取名为学生 t 分布是因为它的发明者 W. S. Gossett(1876 – 1937)是在名为“学生”的刊物上发表了他的论文。Gossett，E. S. Pearson，R. A. Fisher 和 J. Neyman 被称为现代统计学的奠基人。

$$F_X(x)=\int_{-\infty}^{x} f_X(\zeta)\mathrm{d}\zeta \tag{2.5-1}$$

$$P[x_1<X\leqslant x_2]=\int_{x_1}^{x_2} f_X(\zeta)\mathrm{d}\zeta \tag{2.5-2}$$

和

$$P[B]=\int_{\zeta:\ \zeta\in B} f_X(\zeta)\mathrm{d}\zeta \tag{2.5-3}$$

其中$B\in\mathscr{B}$, B是一个事件。式(2.5-3)是基于下列事实得出的：对于连续型随机变量，事件可以写成在R上不相交区间的并，例如，令$B=\{\xi:\ \xi\in\cup_{i=1}^{n}I_i,\ I_iI_j=\phi,\ i\neq j\}$，其中$I_i=(a_i,\ b_j]$，那么

$$\begin{aligned}P[B]&=\int_{a_1}^{b_1} f_X(\zeta)\mathrm{d}\zeta+\int_{a_2}^{b_2} f_X(\zeta)\mathrm{d}\zeta+\cdots+\int_{a_n}^{b_n} f_X(\zeta)\mathrm{d}\zeta\\&=\int_{\zeta:\ \zeta\in B} f_X(\zeta)\mathrm{d}\zeta\end{aligned} \tag{2.5-4}$$

离散型随机变量有阶梯形的分布函数，如图2.5-1所示。

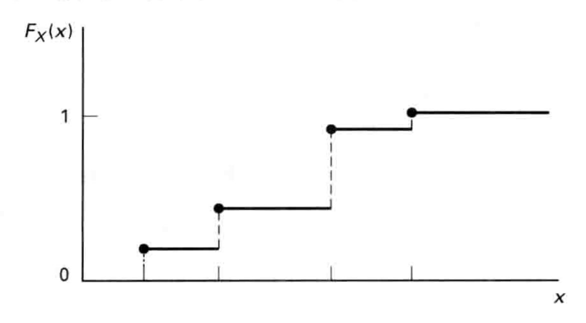


图2.5-1　离散型随机变量的累积分布函数

离散型随机变量的概率测度是概率质量函数[①](PMF)。(离散型)随机变量X的PMF $P_X(x)$定义为

$$\begin{aligned}P_X(x)&=P[X=x]\\&=P[X\leqslant x]-P[X<x]\end{aligned} \tag{2.5-5}$$

因此，在$F_X(x)$的连续处，$P_X(x)=0$，而在$F_X(x)$不连续处，即$F_X(x)$的跳变处，$P_X(x)$有非零值。如果用$F_X(x^-)$表示$P[X<x]$，那么在跳变点x_i, $i=1, 2, \cdots$, $P_X(x_i)$的值可以由$P_X(x_i)=F_X(x_i)-F_X(x_i^-)$计算。

概率质量函数可用在随机实验的结果是可数的时候。确实，$P_X(x_i)$借用了下面的频率解释：做n次实验，令n_i表示x_i出现的次数，那么，对于大的n

$$P_X(x_i)\approx\frac{n_i}{n} \tag{2.5-6}$$

由于PMF与概率的频率表示很接近，所以有时也称为频率函数。

由于离散型随机变量的$F_X(x)$是不连续的，因此，严格地讲，$f_X(x)$是不存在的。然后，

① 概率像质量一样可以保存，概率质量函数中的"质量"就是指这个意思。

如果引入狄拉克δ函数[①]，那么就能够对离散型随机变量的pdf赋值。离散型随机变量的CDF为

$$F_X(x) \triangleq P[X \leqslant x] = \sum_{\text{所有的} x_i \leqslant x} P_X(x_i) \tag{2.5-7}$$

更一般的情况，当X是离散型随机变量时，对于任意事件B

$$P[B] = \sum_{\text{所有的} x_i \in B} P_X(x_i) \tag{2.5-8}$$

常用的离散型随机变量

1. 参数为p的伯努利随机变量$B(0<p<1,\ q\triangleq 1-p)$为

$$P_B(k) = \begin{cases} q, k=0 \\ p, k=1 \\ 0, \text{其他} \end{cases} \tag{2.5-9}$$

$$= q\delta(k) + p\delta(k-1) \qquad \text{利用离散}\ \delta(k)\ \text{函数}^{②} \tag{2.5-10}$$

伯努利随机变量出现在下面的情况：当实验的结果是两种可能的情况之一，例如，在数字序列中的某一位是“1”或者“0”。伯努利PMF可以方便地写成$P_B(k)=p^k q^{1-k}$，其中$k=0$或1，对其他的k，$P_B(k)=0$，对应的CDF为

$$F_B(k) = \begin{cases} 0, k<0 \\ q, k=0 \\ 1, k\geqslant 1 \end{cases}$$

$$= qu(k) + pu(k-1) \ \ \text{利用单位阶跃函数}\ u(k)$$

2. 参数为n和p的二项式随机变量K（$n=1, 2, \cdots; 0<p<1$），k为整数

$$P_K(k) = \begin{cases} \binom{n}{k} p^k q^{n-k}, 0\leqslant k\leqslant n \\ 0, \qquad \text{其他} \end{cases} \tag{2.5-11}$$

$$= \binom{n}{k} p^k q^{n-k} [u(k) - u(n-k)] \tag{2.5-12}$$

二项式随机变量出现在机会游戏、军事防御策略、失效分析等，它对应的CDF为

$$F_K(k) = \begin{cases} 0, & k<0 \\ \sum_{l=0}^{k} \binom{n}{l} p^l q^{n-l}, & 0\leqslant k<n \\ 1, & k\geqslant n \end{cases}$$

其中l, k, n为整数。

3. 参数为$\mu(\mu>0)$的泊松随机变量X，k为整数

$$P_X(k) = \begin{cases} \frac{\mu^k}{k!} e^{-\mu}, 0\leqslant k<\infty \\ 0, \qquad \text{其他} \end{cases} \tag{2.5-13}$$

泊松分布广泛应用于科学和工程的每一个分支(参见1.10节)。泊松PMF用单位阶跃函数$u(k)$可写成

① 有时也称为冲激函数，由英国物理学家Paul A. M. Dirac(1902－1984)而命名，狄拉克函数将在附录B中讨论。

② 当自变量为零时离散δ函数为1，而自变量为其他值时，离散δ函数为0。

$$P_X(k)=\frac{\mu^k}{k!}\mathrm{e}^{-\mu}u(k)$$

其中离散单位阶跃函数定义为

$$u(k)\triangleq\begin{cases}1, & 0\leqslant k<\infty\\0, & -\infty<k<0\end{cases}$$

4. 参数为 $p>0$，$q>0$（$p+q=1$）的几何分布随机变量 K，k 为整数

$$\begin{aligned}P_K(k)&=\begin{cases}pq^k, & 0\leqslant k<\infty\\0, & \text{其他}\end{cases}\\&=pq^ku(k)\end{aligned}\tag{2.5-14}$$

对应的 CDF 用几何级数(参见附录 A)的有限和表示为

$$\begin{aligned}F_K(k)&=\begin{cases}0, & k<0\\p\left(\frac{1-q^{k+1}}{1-q}\right), & 0\leqslant k<\infty\end{cases}\\&=p\left(\frac{1-q^{k+1}}{1-q}\right)u(k)\end{aligned}\tag{2.5-15}$$

这个分布在例 1.9-5 中首次出现，在那里注意变化 pq^{n-1}，$n\geqslant1$，称为几何随机变量[1]。

例 2.5-1 （泊松随机变量的 CDF）下面计算泊松随机变量的 CDF。令 X 为具有参数 μ 的泊松随机变量，由 PMF 的定义，$P_X(k)=\frac{\mu^k}{k!}\mathrm{e}^{-\mu}u(k)$，那么，对于 $k<0$，$F_X(k)=0$，对于 $k\geqslant0$，有

$$\begin{aligned}F_X(k)&=\sum_{l=0}^{k}\frac{\mu^l}{l!}\mathrm{e}^{-\mu}\\&=\left(\sum_{l=0}^{k}\frac{\mu^l}{l!}\right)\mathrm{e}^{-\mu}\end{aligned}$$

表 2.5-1 列出了几种常用离散型随机变量及它们的 PMF 和 CDF。

表 2.5-1 常用离散型随机变量及它们的 PMF 和 CDF 表

类 型	PMF $P_K(k)$	CDF $F_K(k)$
伯努利 p,q	$q\delta(k)+p\delta(k-1)$	$qu(k)+pu(k-1)$
二项式 n,k	$\binom{n}{k}p^kq^{n-k}[u(k)-u(n-k)]$	$\begin{cases}0, & k<0\\\sum_{l=0}^{k}\binom{n}{l}p^lq^{n-l}, & 0\leqslant k<n\\1, & k\geqslant n\end{cases}$
泊松 $\mu>0$	$\frac{\mu k}{k!}\mathrm{e}^{-\mu}u(k)$	$\frac{\gamma(k+1,\mu)}{k!}$[2] $\times u(k)$
几何 p,q	$pq^ku(k)$	$p\left(\frac{1-q^{k+1}}{1-q}\right)u(k)$

① 注意，我们有时说一般意义的概率分布，并非都是指分布函数，这里用 PMF 来说明几何分布。

② 不完备伽马函数的定义参见附录 B。

有时候随机变量既不是纯离散的、也不是纯连续的，我们称这样的随机变量为混合型随机变量，混合型随机变量的 CDF 如图 2.5-2 所示，因此，$F_X(x)$是不连续的，但不是阶梯函数。

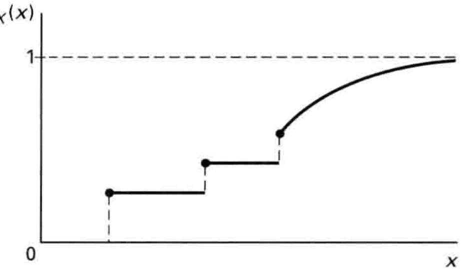


图 2.5-2　混合型随机变量的 CDF

连续和离散随机变量的区别是人为的，连续和离散型随机变量常常看成不同的目标，尽管它们真正的区别在于前者的 CDF 是连续的，而后者则不是。引入δ函数以后，就可以把它们当成一类来处理，对连续和离散型随机变量概率的计算只需要对 pdf 积分即可。

现在回到式(2.5-7)，对于离散型随机变量，可以写为

$$F_X(x)=\sum_{i=-\infty}^{\infty}P_X(x_i)u(x-x_i) \tag{2.5-16}$$

利用附录 B 中的δ函数，得

$$f_X(x)=\frac{\mathrm{d}F_X(x)}{\mathrm{d}_X}=\sum_{i=-\infty}^{\infty}P_X(x_i)\delta(x-x_i) \tag{2.5-17}$$

例 2.5-2　(计算实例)令X是离散型随机变量，分布函数如图 2.5-3(a)所示，X的 pdf 为

$$f_X(x)=\frac{\mathrm{d}F_X}{\mathrm{d}x}=0.2\delta(x)+0.6\delta(x-1)+0.2\delta(x-3)$$

图形如图 2.5-3(b)所示。为了计算离散型随机变量的概率，积分区间必须仔细选择，因此，

$$F_X(x)=\int_{-\infty}^{x^+}f_X(\zeta)\mathrm{d}\zeta$$

在有δ函数的地方，积分就包含了这个δ函数。

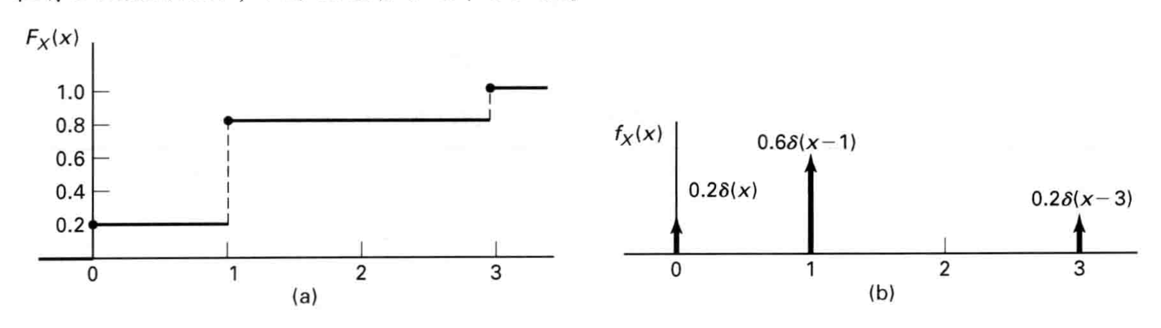


图 2.5-3　(a) 离散随机变量的 CDF；(b) 使用δ函数的X的 pdf

类似地，$P[x_1<X\leqslant x_2]$包含了区间

x_1　　　x_2

这就包含了x_2处的冲激(如果在那里有一个冲激的话)，但排除了x_1处的冲激。另一方面，$P[x_1\leqslant X<x_2]$包含了区间



因此

$$P(x_1 \leqslant X < x_2) = \int_{x_1^-}^{x_2^-} f_X(\zeta)\mathrm{d}\zeta$$

应用到本例中，由这些公式得出

$$P[X \leqslant 1.5] = F_X(1.5) = 0.8 \tag{2.5-18}$$

$$P[1 < X \leqslant 3] = 0.2 \tag{2.5-19}$$

$$P[1 \leqslant X < 3] = 0.6 \tag{2.5-20}$$

例2.5-3 （计算实例）参数为 a 的泊松分布的pdf为

$$f_X(x) = \mathrm{e}^{-a}\sum_{k=0}^{\infty}\frac{a^k}{k!}\delta(x-k)$$

例2.5-4 （计算实例）二项式分布 $b(k;\ n,\ p)$ 的pdf

$$f_X(x) = \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\delta(x-k)$$

例2.5-5 （计算实例）如图2.5-4所示的混合型随机变量的pdf。(1)常数 K 是多少？(2)计算 $P[X \leqslant 5]$，$P[5 \leqslant X < 10]$；(3)画出分布函数。

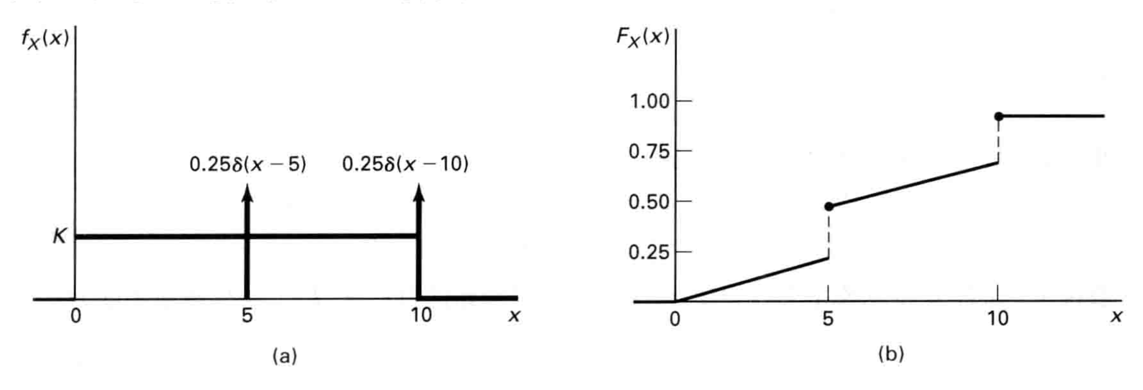


图2.5-4 (a) 例2.5-5的混合型随机变量的pdf；(b)计算出的pdf

解

(1)由于

$$\int_{-\infty}^{\infty} f_X(\zeta)\mathrm{d}\zeta = 1$$

可得到 $10K + 0.25 + 0.25 = 1 \Rightarrow K = 0.05$。

(2)由于 $P[X \leqslant 5] = P[X < 5] + P[X = 5]$，在 $x = 5$ 处有一个冲激必须包括在内，因此

$$P[X \leqslant 5] = \int_0^{5^+}[0.05 + 0.25\delta(\zeta - 5)]\mathrm{d}\zeta$$
$$= 0.5$$

为了计算 $P[5\leqslant X<10]$，我们排除 $x=10$ 处的冲激，但包含 $x=5$ 处的冲激，因此

$$P[5 \leqslant X < 10] = \int_{5^-}^{10^-} [0.05 + 0.25\delta(\zeta - 5)]\mathrm{d}\zeta$$
$$= 0.5$$

(3) 对图 2.5-4(a) 的 pdf 积分可得到 X 的概率分布函数，图形画在图 2.5-4(b) 中。

2.6 条件和联合分布函数和概率密度函数

考虑一个事件 C，它是由所有那些满足 $X(\zeta)\leqslant x$ 且 $\zeta\in B\subset\Omega$ 的实验结果 $\zeta\in\Omega$ 组成，其中 B 是另一个事件。那么根据定义，事件 C 是两个事件 $\{\zeta: X(\zeta)\leqslant x\}$ 和 $\{\zeta:\zeta\in B\}$ 的交集。我们定义在事件 B 给定的条件下 X 的条件分布函数为

$$F_X(x|B) \triangleq \frac{P[C]}{P[B]} = \frac{P[X \leqslant x, B]}{P[B]} \tag{2.6-1}$$

其中 $P[X\leqslant x, B]$ 是联合事件 $\{X(\zeta)\leqslant x\}\cap B$ 和 $(P[B]\neq 0)$ 的概率。如果 $x=\infty$，事件 $\{X\leqslant\infty\}$ 是必然事件 Ω，由于 $\Omega\cap B=B$，所以 $F_X(\infty|B)=1$。类似地，如果 $x=-\infty$，$\{X\leqslant -\infty\}=\phi$，由于 $\Omega\cap\phi=\phi$，所以 $F_X(-\infty|B)=0$。按这种方式继续，不难证明 $F_X(x|B)$ 有一般分布函数的所有性质，例如 $x_1\leqslant x_2\rightarrow F_X(x_1|B)\leqslant F_X(x_2|B)$。

例如，考虑事件 $\{X\leqslant x_2, B\}$，写成（假定 $x_2\geqslant x_1$）

$$\{X \leqslant x_2, B\} = \{X \leqslant x_1, B\} \cup \{x_1 < X \leqslant x_2, B\}$$

由于右边的两个事件是不相交的，它们的概率相加，于是

$$P[X \leqslant x_2, B] = P[X \leqslant x_1, B] + P[x_1 < X \leqslant x_2, B]$$

或者

$$P[X \leqslant x_2|B]P[B] = P[X \leqslant x_1|B]P[B] + P[x_1 < X \leqslant x_2|B]P[B]$$

因此当 $P[B]\neq 0$ 时，经整理后我们得到

$$\begin{aligned} P[x_1 < X \leqslant x_2|B] &= P[X \leqslant x_2|B] - P[X \leqslant x_1|B] \\ &= F_X(x_2|B) - F_X(x_1|B) \end{aligned} \tag{2.6-2}$$

一般说来，事件 B 依据概率空间 $\{R, \mathscr{B}, P_X\}$ 而不是原始的 $\{\Omega, \mathscr{F}, P\}$，条件概率密度为

$$f_X(x|B) \triangleq \frac{\mathrm{d}F_X(x|B)}{\mathrm{d}x} \tag{2.6-3}$$

下面给出一些例子。

例 2.6-1 （条件 CDF 的计算）令 $B\triangleq\{X\leqslant 10\}$，计算 $F_X(x|B)$。

(1) 对于 $x\geqslant 10$，事件 $\{x\leqslant 10\}$ 是事件 $\{X\leqslant x\}$ 的子集，因此，$P[X\leqslant 10, X\leqslant x]=P[X\leqslant 10]$，利用式(2.6-1)，得

$$F_X(x|B) = \frac{P[X \leqslant x, X \leqslant 10]}{P[X \leqslant 10]} = 1$$

(2) 对于 $x\leqslant 10$，事件 $\{X\leqslant x\}$ 是事件 $\{x\leqslant 10\}$ 的子集，因此，$P[X\leqslant 10, X\leqslant x]=P[X\leqslant x]$，且

$$F_X(x|B) = \frac{P[X \leqslant x]}{P[X \leqslant 10]}$$

其结果画在图2.6-1中。当 $B=\{b<X\leqslant a\}$ 时，$F_X(x|B)$ 的计算留给读者练习。

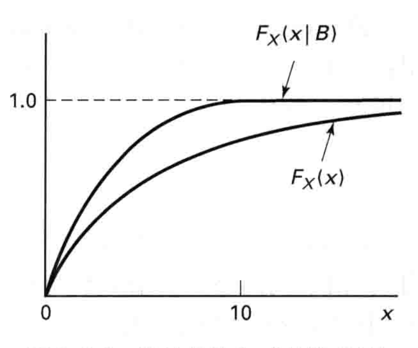


图2.6-1 X 的条件和无条件CDF

例2.6-2 （泊松条件CDF）令 X 是一个参数为 $\mu\ (>0)$ 的泊松随机变量，我们希望计算在给定事件 $\{X=0, 2, 4, \cdots\}\triangleq\{X\text{为偶数}\}$ 的条件下 X 的条件PMF和条件CDF。首先观察到 $P[X\text{为偶数}]$ 为

$$P[X=0,2,\cdots]=\sum_{k=0,2,\cdots}^{\infty}\frac{\mu^k}{k!}\mathrm{e}^{-\mu}$$

那么，当 X 为奇数时，我们有

$$P[X=1,3,\cdots]=\sum_{k=1,3,\cdots}^{\infty}\frac{\mu^k}{k!}\mathrm{e}^{-\mu}$$

由这些关系，得

$$\begin{aligned}\sum_{k\geqslant 0\text{且}k\text{为偶数}}\frac{\mu^k}{k!}\mathrm{e}^{-\mu}-\sum_{k\geqslant 0\text{且}k\text{为奇数}}\frac{\mu^k}{k!}\mathrm{e}^{-\mu}&=\sum_{k=0}^{\infty}\frac{\mu^k}{k!}(-1)^k\mathrm{e}^{-\mu}\\&=\sum_{k=0}^{\infty}\frac{(-\mu)^k}{k!}\mathrm{e}^{-\mu}\\&=\mathrm{e}^{-\mu}\mathrm{e}^{-\mu}\\&=\mathrm{e}^{-2\mu}\end{aligned}$$

和

$$\sum_{k\geqslant 0\text{且}k\text{为偶数}}\frac{\mu^k}{k!}\mathrm{e}^{-\mu}+\sum_{k\geqslant 0\text{且}k\text{为奇数}}\frac{\mu^k}{k!}\mathrm{e}^{-\mu}=1$$

因此，$P[X\text{为偶数}]=P[X=0, 2, \cdots]=\frac{1}{2}(1+\mathrm{e}^{-2\mu})$。利用条件PMF的定义，我们有

$$P_X(k|X\text{为偶数})=\frac{P[X=k, X\text{为偶数}]}{P[X\text{为偶数}]}$$

如果 k 为偶数，那么 $\{X=k\}$ 是 $\{X\text{为偶数}\}$ 的子集。如果 k 为奇数，$\{X=k\}\cap\{X\text{为偶数}\}=\phi$，因此，当 k 为偶数时，$P[X=k, X\text{为偶数}]=P[X=k]$，当 k 为奇数时，$P[X=k, X\text{为偶数}]$ 为零。所以，我们有

$$P_X(k|X\text{为偶数})=\begin{cases}\frac{2}{(1+2\mathrm{e}^{-\mu})}\frac{\mu^k}{k!}\mathrm{e}^{-\mu}, & k\geqslant 0\text{且为偶数}\\ 0, & k\text{为奇数}\end{cases}$$

那么，条件CDF为

$$\begin{aligned}F_X(x|X\text{为偶数})&=\sum_{\text{所有}k\leqslant x}P_X(k|X\text{为偶数})\\&=\sum_{\substack{0\leqslant k\leqslant x\\ \text{且}k\text{为偶数}}}\frac{2}{(1+2\mathrm{e}^{-\mu})}\frac{\mu^k}{k!}\mathrm{e}^{-\mu}\end{aligned}$$

下面推导一些包含条件CDF和条件pdf的重要公式。

分布函数表示为条件分布函数的加权和　在式(1.6-7)给出了事件 B 的概率用 n 个互不相容的事件$\{\triangleq A_i\}$表示，其中 $i=1, 2, \cdots, n$，这些事件$\{A_i\}$是定义在与事件 B 相同的概率空间上。用 $B=\{X\leqslant x\}$，由式(1.6-7)可立即得出

$$F_X(x)=\sum_{i=1}^{n}F_X(x|A_i)P[A_i] \tag{2.6-4}$$

式(2.6-4)将 $F_X(x)$ 描述为条件分布函数的加权和。式(2.6-4)可以看成对条件 CDF 进行平均[①]。由于我们还没有给出平均的具体表示(将在第 4 章讨论)，这里只要求读者接受这一命名，因为在技术文献中很有用。

例 2.6-3　(有瑕疵的存储芯片)在计算机存储芯片的自动制造中，公司 Z 每生产 5 片好的芯片就会有 1 片有瑕疵的芯片，有瑕疵的芯片(DC)失效的时间服从 CDF

$$F_X(x|\text{DC})=(1-\mathrm{e}^{-x/2})u(x)\quad (x\text{的单位为月})$$

而好的芯片(GC)失效的时间服从 CDF

$$F_X(x|\text{GC})=(1-\mathrm{e}^{-x/10})u(x)\quad (x\text{的单位为月})$$

芯片的好坏是看不出来的。芯片卖出去了，在使用 6 个月之前失效的概率是多少？

解　由式(2.6-4)，芯片的无条件 CDF 为

$$F_X(x)=F_X(x|\text{DC})P[\text{DC}]+F_X(x|\text{GC})P[\text{GC}]$$

其中，$P[\text{DC}]$和$P[\text{GC}]$分别是选择一片有瑕疵的芯片和一片好的芯片的概率，根据给定的数据，$P[\text{DC}]=1/6$，$P[\text{GC}]=5/6$，因此

$$\begin{aligned}F_X(6)&=[1-\mathrm{e}^{-3}]\tfrac{1}{6}+[1-\mathrm{e}^{-0.6}]\tfrac{5}{6}\\&=0.158+0.376=0.534\end{aligned}$$

概率密度函数的贝叶斯公式　考虑定义在同一个概率空间上的事件 B 和$\{X=x\}$，由条件概率的定义，写成下式似乎是合理的

$$P[B|X=x]=\frac{P[B,X=x]}{P[X=x]} \tag{2.6-5}$$

式(2.6-5)的问题是：如果 X 是连续型随机变量，那么 $P[X=x]=0$，因此，式(2.6-5)没有定义。然而，我们计算 $P[B|X=x]$可以通过对包含事件$\{x<X\leqslant x+\Delta x\}$的概率取极限来得到。考虑表达式

$$P[B|x<X\leqslant x+\Delta x]=\frac{P[x<X\leqslant x+\Delta x|B]P[B]}{P[x<X\leqslant x+\Delta x]}$$

如果(i)在表达式右边的分母、分子除以 Δx，(ii)利用这一事实：$P[x<X\leqslant x+\Delta x|B]=F(x+\Delta x|B)-F(x|B)$，(iii)取 $\Delta x\to 0$ 时的极限，可得到

$$\begin{aligned}P[B|X=x]&=\lim_{\Delta x\to 0}P[B|x<X\leqslant x+\Delta x]\\&=\frac{f_X(x|B)P[B]}{f_X(x)},\qquad f(x)\neq 0\end{aligned} \tag{2.6-6}$$

式(2.6-6)称为在 $X=x$ 的条件下事件 B 的后验概率(或后验密度)，在式(2.6-6)的两边乘以

---

① 由于这一原因，当 $F_X(x)$ 用式(2.6-4)表示时，有时也称为平均分布函数。

$f_X(x)$并对x积分，得出如下重要的结果：

$$P[B]=\int_{-\infty}^{\infty}P[B|X=x]f_X(x)\mathrm{d}x \tag{2.6-7}$$

与本节采用的符号一致，$P[B]$有时也称为B的平均概率，这是由式(2.6-7)的形式得到的。

例2.6-4　(检测开关的闭合)一个信号X可能来自三个不同的源，分别标以A、B和C。来自A的信号服从$N(-1,4)$，来自B的信号服从$N(0,1)$，来自C的信号服从$N(1,4)$，为了让信号到达终点R，线上的开关必须闭合。当在R观察到信号X时，只有一个开关是闭合的，但不知道是哪个开关闭合，然而知道开关a闭合的次数是开关b的两倍，开关b闭合的次数是开关c的两倍，如图2.6-2所示。

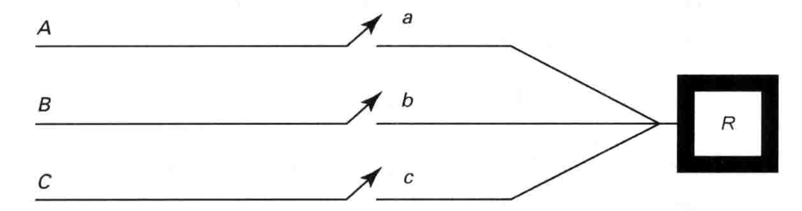


图2.6-2　根据观察到的信号，接收机R必须决定哪个开关是闭合的，或等价于源A、B、C中的哪一个对信号负责。接收和开机时只有一个开关闭合

(1)计算$P[X\leqslant -1]$。

(2)假定观测到事件$\{X>-1\}$，这个信号最有可能来自哪个源？

解　(a)令$P[A]$表示在R得到的观测是源于A的概率，即开关a闭合的概率。$P[B]$和$P[C]$与此类似。那么，由题目中开关的信息，我们有$P[A]=2P[B]=4P[C]$，且$P[A]+P[B]+P[C\}=1$。因此，$P[A]=4/7$，$P[B]=2/7$，$P[C]=1/7$。接下来计算$P[X\leqslant -1]$，

$$P[X\leqslant -1]=P[X\leqslant -1|A]P[A]+P[X\leqslant -1|B]P[B]+P[X\leqslant -1|C]P[C]$$

其中

$$P[X\leqslant -1|A]=1/2 \tag{2.6-8}$$

$$P[X\leqslant -1|B]=1/2-\mathrm{erf}(1)=0.159 \tag{2.6-9}$$

$$P[X\leqslant -1|C]=1/2-\mathrm{erf}(1)=0.159 \tag{2.6-10}$$

因此$P[X\leqslant -1]=1/2\times 4/7+0.159\times 2/7+0.159\times 1/7\approx 0.354$。

(b)计算$\max\{P[A|X>-1],P[B|X>-1],P[C|X>-1]\}$，注意到$P[X>-1|A]=1-P[X\leqslant -1|A]$，对$B$和$C$也有类似关系。对于源$A$，利用贝叶斯公式，我们有

$$P[A|X>-1]=\frac{\{1-P[X\leqslant -1|A]\}\times P[A]}{1-P[X\leqslant -1]}$$

将前面计算的结果代入，得$P[A|X>-1]=0.44$。

对其他的源，重复以上的计算过程，得

$$P[B|X>-1]=0.372 \tag{2.6-11}$$

$$P[C|X>-1]=0.186 \tag{2.6-12}$$

因此，按后验概率最大的原则，源A是事件$\{X>-1\}$最有可能的原因。

泊松变换 式(2.6-7)一个重要的特例是所谓的泊松变换，其中B是随机变量Y从集$\{0,1,\cdots\}$取整数值的事件，即$B\triangleq\{Y=k\}$，X是泊松参数，这里把该参数看成一个随机变量，其概率密度为$f_X(x)$，一般的泊松分布为

$$P[Y=k]=\mathrm{e}^{-\mu}\frac{\mu^k}{k!},\qquad k\geqslant 0 \tag{2.6-13}$$

其中μ是在给定区间上(时间、距离、体积等)事件的平均数，参数μ看成常数。但在许多情况下，所考虑的物理现象决定了μ自身也是随机的，必须把它看成随机实验的结果。这样就有两个随机因素，μ的随机值和随机结果$\{Y=k\}$，当μ是随机的时候，用一个随机变量X去取代它更合适。因此，对于给定的随机实验结果$\{X=x\}$，概率$P[Y=k|X=x]$是泊松的，但事件$\{Y=k\}$的无条件概率不必是泊松的。因为事件数和泊松参数都是随机的，这种情况有时也称为双随机。由式(2.6-7)，我们得到Y的无条件 PMF 为

$$P_Y(k)=\int_0^{\infty}\frac{x^k}{k!}\mathrm{e}^{-x}f_X(x)\mathrm{d}x,\qquad k\geqslant 0 \tag{2.6-14}$$

上式称为泊松变换，如果$P_Y(k)$能够由实验得到，那么就可以利用上式得到$f_X(x)$。由$P_Y(k)$得到$f_X(x)$的机制称为逆泊松变换①，推导如下，令

$$F(\omega)\triangleq\frac{1}{2\pi}\int_0^{\infty}\mathrm{e}^{\mathrm{j}\omega x}\mathrm{e}^{-x}f_X(x)\mathrm{d}x \tag{2.6-15}$$

即$\mathrm{e}^{-x}f_X(x)$的傅里叶逆变换，由于

$$\mathrm{e}^{\mathrm{j}\omega x}=\sum_{k=0}^{\infty}[\mathrm{j}\omega x]^k/k! \tag{2.6-16}$$

可得到

$$\begin{aligned}F(\omega)&=\frac{1}{2\pi}\sum_{k=0}^{\infty}(\mathrm{j}\omega)^k\int_0^{\infty}\frac{x^k}{k!}\mathrm{e}^{-x}f_X(x)\mathrm{d}x\\&=\frac{1}{2\pi}\sum_{k=0}^{\infty}\mathrm{j}\omega^kP_Y(k)\end{aligned} \tag{2.6-17}$$

因此，如果$P_Y(k)$已知的话，$F(\omega)$也就知道了。取$F(\omega)$的傅里叶正变换，得

$$\mathrm{e}^{-x}f_X(\omega)=\mathrm{e}^{x}\int_{-\infty}^{\infty}F(\omega)\mathrm{e}^{-\mathrm{j}\omega x}\mathrm{d}\omega$$

即

$$f_X(x)=\mathrm{e}^{x}\int_{-\infty}^{\infty}F(\omega)\mathrm{e}^{-\mathrm{j}\omega x}\mathrm{d}\omega \tag{2.6-18}$$

式(2.6-18)是我们要找的逆关系，总结如下：如果知道$P_Y(k)$，就可以计算$F(\omega)$，一旦知道了$F(\omega)$就可以通过傅里叶变换得到$f_X(x)$。下面通过一个光通信的应用来说明泊松变换。

例 2.6-5 (光通信)在光通信系统中，来自发射器的光打到光检测器上会产生光电流，光电流是由价电子变成的传导电子组成(如图 2.6-3 所示)。

由物理学知道，如果发射器使用恒定强度的相干激光，激光强度即泊松参数X的 pdf 为

$$f_X(x)=\delta(x-x_0),\qquad x_0>0 \tag{2.6-19}$$

① 注意，$F(\omega)$不是 CDF。

其中 x_0 是激光强度，它只是一个常数。另一方面，如果发射器采用热照射，那么，泊松参数 X 服从指数分布

$$f_X(x)=\frac{1}{\mu}\mathrm{e}^{-(1/\mu)x}u(x) \tag{2.6-20}$$

其中 $\mu>0$ 刚好是一个参数，但是后面将要证明这个参数是 X 的均值。计算电子计数变量 Y 的 PMF。

解 对于相干激光照射，由式(2.6-14)

$$P_Y(k)=\int_0^{\infty}\frac{x^k}{k!}\mathrm{e}^{-x}\delta(x-x_0)\mathrm{d}x \tag{2.6-21}$$

$$=\frac{x_0^k}{k!}\mathrm{e}^{-x_0},k\geqslant 0 \tag{2.6-22}$$

因此，对于相干激光照射，光电子服从泊松分布。对于热照射，我们有

$$\begin{aligned}P_Y(k)&=\int_0^{\infty}\frac{x^k}{k!}\mathrm{e}^{-x}\frac{1}{\mu}\mathrm{e}^{-x/\mu}\mathrm{d}x\\&=\frac{1}{\mu k!}\int_0^{\infty}x^k\mathrm{e}^{-x/\alpha}\mathrm{d}x,\text{ 其中 }\alpha\triangleq\frac{\mu}{\mu+1}\\&=\frac{\alpha^{k-1}}{\mu k!}\int_0^{\infty}z^k\mathrm{e}^{-z}\mathrm{d}z,\text{ 变量替换 }z\triangleq x/\alpha\\&=\frac{\alpha^{k-1}}{\mu k!}\Gamma(k+1),\quad\text{其中 }\Gamma\text{ 表示伽马函数(参见附录B)}\\&=\frac{\alpha^{k-1}}{\mu k!}k!\\&=\frac{\alpha^{k-1}}{\mu}\\&=\frac{\mu^k}{[1+\mu]^{k+1}},k\geqslant 0\end{aligned} \tag{2.6-23}$$

这个 PMF 分布称为几何分布，有时也称为玻色–爱因斯坦(Bose-Einstein)统计[2-4]，它服从有趣的递推关系

$$P_Y(k+1)=\frac{\mu}{1+\mu}P_Y(k) \tag{2.6-24}$$

采用不同的照射方式，光电流会有相当的不同。

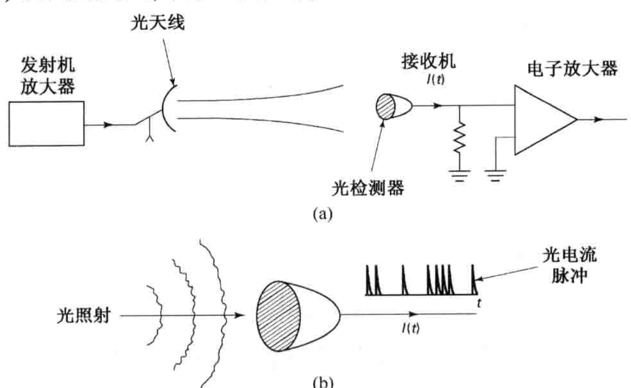


图 2.6-3 (a)光通信系统；(b)光检测器的输出电流

联合概率分布函数与联合概率密度 正如2.1 节所述，在概率空间上有可能定义多个随机变量，例如，考虑一个概率空间$(\Omega, \mathscr{F}, P)$，考虑的随机实验是由同时投掷两枚均匀硬币组成，这里顺序并不重要，基本实验结果为ζ_1 = HH、ζ_2 = HT、ζ_3 = TT，样本空间为Ω = {HH, HT, TT}事件的σ 域为ϕ、Ω、{HT}、{TT}、{HH}、{TT 或 HT}、{HH 或 HT}、{HH 或 TT}，概率很容易计算，分别为0、1、1/2、1/4、1/4、3/4、3/4 和1/2。现在定义两个随机变量

$$X_1(\zeta) = \begin{cases} 0, & \text{如果至少有一个正面}H \\ 1, & \text{其他} \end{cases} \tag{2.6-25}$$

$$X_2(\zeta) = \begin{cases} -1, & \text{如果有一个正面}H\text{和一个反面}T \\ +1, & \text{其他} \end{cases} \tag{2.6-26}$$

那么，$P[X_1=0]=3/4$，$P[X_1=1]=1/4$，$P[X_2=-1]=1/2$，$P[X_2=1]=1/2$，另外，我们也很容易计算联合事件的概率，例如，$P[X_1=0, X_2=1]=P[\{\text{HH}\}]=1/4$。

在概率空间上定义一个以上的随机变量时，有可能定义了一个退化了的随机变量。例如，考虑一个随机实验是由转轮停下来指向的数字ζ组成(0 ~ 100)，假定我们令$X_1(\zeta)=\zeta$，$X_2(\zeta)=e^{\zeta}$，这种情况就是退化，因为观察到一个随机变量以后，也就完全指定了另一个。实际上，不确定性只是与一个随机变量有关，并不是两个，我们不妨忘记观察另一个。如果在概率空间上定义一个以上的随机变量，如果随机实验足够复杂或者实验结果足够丰富，退化就可以避免。例如，在开始我们考虑的例子中，观察到$X_1=0$，并不指定X_2的值，而观察到$X_2=1$时并不指定X_1的值。

事件$\{X\leqslant x, Y\leqslant y\} \triangleq \{X\leqslant x\}\cap\{Y\leqslant y\}$是由所有$\zeta\in\Omega$且满足$X(\zeta)\leqslant x$和$Y(\zeta)\leqslant y$的随机实验结果组成。由事件$\{X\leqslant x, Y\leqslant y\}$导出的点集在$x'y'$平面上是图 2.6-4 所示的阴影区域，在图中$x$、$y$表示的数是正的，一般情况下它们可能是任意的值。$X$和$Y$的联合累积分布函数定义为

$$F_{XY}(x,y)=P[X\leqslant x, Y\leqslant y] \tag{2.6-27}$$

根据定义，$F_{XY}(x, y)$是一个概率，因此，对所有的x和y，$F_{XY}(x, y)\geqslant 0$。由于$\{X\leqslant\infty, Y\leqslant\infty\}$是一个必然事件，所以$F_{XY}(\infty, \infty)=1$。与必然事件有关的点集是整个$x'y'$平面，事件$\{X\leqslant-\infty, Y\leqslant-\infty\}$是不可能事件，因此，$F_{XY}(-\infty, -\infty)=0$。读者可以考虑一下事件$\{X\leqslant x, Y\leqslant-\infty\}$和$\{X\leqslant-\infty, Y\leqslant y\}$，它们也是不可能事件吗?

由于$\{X\leqslant\infty\}$和$\{Y\leqslant\infty\}$是必然事件，对于任意事件B，$B\cap\Omega=B$，于是得到

$$\begin{aligned}\{X\leqslant x, Y\leqslant\infty\} &= \{X\leqslant x\}\cap\{Y\leqslant\infty\} \\ &= \{X\leqslant x\}\cap\Omega \\ &= \{X\leqslant x\}\end{aligned} \tag{2.6-28}$$

因此

$$F_{XY}(x,\infty)=F_X(x) \tag{2.6-29a}$$

$$F_{XY}(\infty,y)=F_Y(y) \tag{2.6-29b}$$

图 2.6-4 与事件$\{X\leqslant x, Y\leqslant y\}$有关的点集

如果$F_{XY}(x, y)$是连续的、可导的，联合 pdf 可以由下式得到：

$$f_{XY}(x,y)=\frac{\partial^2}{\partial x\,\partial y}[F_{XY}(x,y)] \tag{2.6-30}$$

很显然

$$f_{XY}(x,y)\mathrm{d}x\,\mathrm{d}y=P[x<X\leqslant x+\mathrm{d}x,y<Y\leqslant y+\mathrm{d}y]$$

因此，对于所有(x, y)，$f_{XY}(x, y)\geqslant 0$。

对式(2.6-30)两次积分，得

$$F_{XY}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_{XY}(\zeta,\eta)\mathrm{d}\eta\,\mathrm{d}\zeta \tag{2.6-31}$$

式(2.6-31)表明$F_{XY}(x, y)$是对非负函数$f_{XY}(x, y)$在图2.6-4阴影部分上的积分。很显然，对$f_{XY}(x, y)$积分的区域越大，概率也越大。由此得出一个很明显但很重要的结论：如果(x_1, y_1)、(x_2, y_2)表示两对数，如果$x_1\leqslant x_2$、$y_1\leqslant y_2$，那么，$F_{XY}(x_1, y_1)\leqslant F_{XY}(x_2, y_2)$。一般说来，$F_{XY}(x, y)$随$(x, y)$向上和向右的移动而减小，随$(x, y)$向下和向左的移动而增加。此外，$F_{XY}(x, y)$是从上和从右连续的，即在不连续点，如$x_0$、$y_0$，以及$\varepsilon>0$、$\delta>0$

$$F_{XY}(x_0,y_0)=\lim_{\substack{\varepsilon\to 0\\ \delta\to 0}}F_{XY}(x_0+\varepsilon,y_0+\delta)$$

因此，在不连续点，$F_{XY}(x, y)$假定了取右边和上边的值。

联合 CDF $F_{XY}(x, y)$的性质

(i) $F_{XY}(\infty, \infty)=1$，$f_{XY}=(-\infty, y)=F_{XY}(x, -\infty)=0$，此外，$F_{XY}(x, \infty)=F_X(x)$；$F_{XY}(\infty, y)=F_Y(y)$。

(ii) 如果$x_1\leqslant x_2$，$y_1\leqslant y_2$，那么$F_{XY}(x_1, y_1)\leqslant F_{XY}(x_2, y_2)$。

(iii) $F_{XY}(x, y)=\lim\limits_{\substack{\varepsilon\to 0\\ \delta\to 0}}F_{XY}(x+\varepsilon, y+\delta)$　$\varepsilon,\delta>0$(从右和从上连续)。

(iv)对于所以$x_2\geqslant x_1$和$y_2\geqslant y_1$，我们有

$$F_{XY}(x_2,y_2)-F_{XY}(x_2,y_1)-F_{XY}(x_1,y_2)+F_{XY}(x_1,y_1)\geqslant 0$$

最后也是最关键的性质(iv)是一维CDF不减性质在二维的推广，一维CDF不减性质是：对所有$x_2\geqslant x_1$，$F_X(x_2)-F_X(x_1)\geqslant 0$。这也是事件$\{x_1<X\leqslant x_2, y_1<Y\leqslant y_2\}$具有非负概率的需要，由这个事件导出的点集如图2.6-5所示。

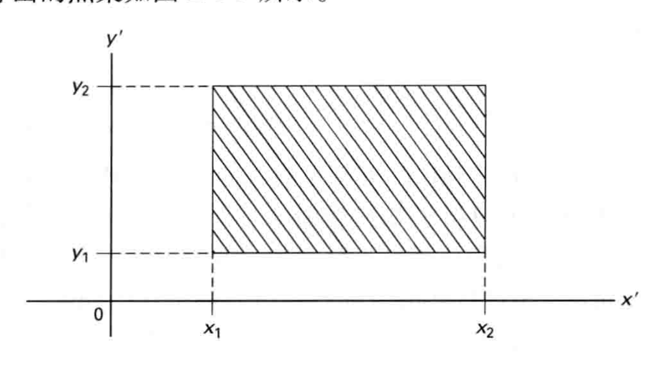


图2.6-5　事件$\{x_1<X\leqslant x_2, y_1<Y\leqslant y_2\}$的点集

这一概率计算的关键是要观察到将集$\{X\leqslant x_2, Y\leqslant y_2\}$分解成下列不相交集：

$$\begin{aligned}\{X\leqslant x_2,Y\leqslant y_2\} &= \{x_1<X\leqslant x_2,y_1<Y\leqslant y_2\}\\ &\cup\{x_1<X\leqslant x_2,Y\leqslant y_1\}\cup\{X\leqslant x_1,y_1<Y\leqslant y_2\}\\ &\cup\{X\leqslant x_1,Y\leqslant y_1\}\end{aligned} \tag{2.6-32}$$

现在利用由公理1[参见式(1.5-3)]导出的结果，得

$$\begin{aligned}F_{XY}(x_2,y_2) &= P[x_1<X\leqslant x_2,y_1<Y\leqslant y_2]\\ &+P[x_1<X\leqslant x_2,Y\leqslant y_1]+P[X\leqslant x_1,y_1<Y\leqslant y_2]\\ &+F_{XY}(x_1,y_1)\end{aligned} \tag{2.6-33}$$

根据定积分的基本性质，式(2.6-33)右边的第二项、第三项能够分别写成

$$\begin{aligned}\int_{x_1}^{x_2}\int_{-\infty}^{y_1}f_{XY}(\zeta,\eta)\mathrm{d}\zeta\,\mathrm{d}\eta &= \int_{-\infty}^{x_2}\int_{-\infty}^{y_1}f_{XY}(\zeta,\eta)\mathrm{d}\zeta\,\mathrm{d}\eta\\ &-\int_{-\infty}^{x_1}\int_{-\infty}^{y_1}f_{XY}(\zeta,\eta)\mathrm{d}\zeta\,\mathrm{d}\eta\end{aligned} \tag{2.6-34}$$

$$\begin{aligned}\int_{-\infty}^{x_1}\int_{y_1}^{y_2}f_{XY}(\zeta,\eta)\mathrm{d}\zeta\,\mathrm{d}\eta &= \int_{-\infty}^{x_1}\int_{-\infty}^{y_2}f_{XY}(\zeta,\eta)\mathrm{d}\zeta\,\mathrm{d}\eta\\ &-\int_{-\infty}^{x_1}\int_{-\infty}^{y_1}f_{XY}(\zeta,\eta)\mathrm{d}\zeta\,\mathrm{d}\eta\end{aligned} \tag{2.6-35}$$

而这些等式右边的项都是分布的，因此，式(2.6-34)和式(2.6-35)变成了

$$\int_{x_1}^{x_2}\int_{-\infty}^{y_1}f_{XY}(\zeta,\eta)\mathrm{d}\zeta\,\mathrm{d}\eta = F_{XY}(x_2,y_1)-F_{XY}(x_1,y_1) \tag{2.6-36}$$

$$\int_{-\infty}^{x_1}\int_{y_1}^{y_2}f_{XY}(\zeta,\eta)\mathrm{d}\zeta\,\mathrm{d}\eta = F_{XY}(x_1,y_2)-F_{XY}(x_1,y_1) \tag{2.6-37}$$

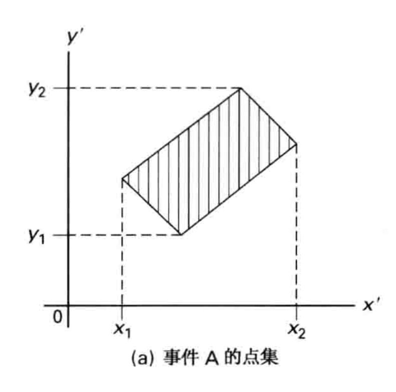
再回到式(2.6-33)，利用式(2.6-36)和式(2.6-37)，得

$$\begin{aligned}F_{XY}(x_2,y_2) &= P[x_1<X\leqslant x_2,y_1<Y\leqslant y_2]\\ &+F_{XY}(x_2,y_1)-F_{XY}(x_1,y_1)+F_{XY}(x_1,y_2)-F_{XY}(x_1,y_1)\\ &+F_{XY}(x_1,y_1)\end{aligned} \tag{2.6-38}$$

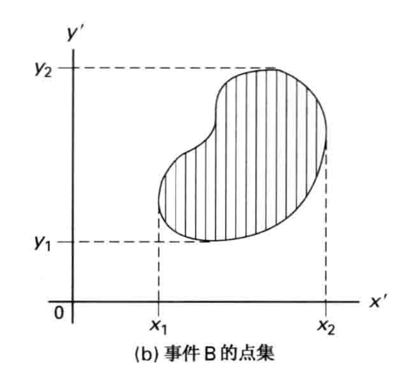
经过化减和整理后，将希望的量出现在左边，最终得到

$$\begin{aligned}P[x_1<X\leqslant x_2,y_1<Y\leqslant y_2] &= F_{XY}(x_2,y_2)-F_{XY}(x_2,y_1)\\ &-F_{XY}(x_1,y_2)+F_{XY}(x_1,y_1)\end{aligned} \tag{2.6-39}$$

式(2.6-39)对任何独立与否的随机变量X、Y都是成立的。在应用式(2.6-39)时必须小心。例如，图2.6-6(a)和(b)中画出了包括了随机变量X、Y的两个区域A和$B\{x_1<X\leqslant x_2\}$和$\{y_1<Y\leqslant y_2\}$。但在式(2.6-39)这里应用是不合适的，因为没有一个矩形区域的边与轴平行。在图2.6-6(a)所表示的情况，通过旋转的坐标变换可以解决这个问题，但这涉及随机变量变换的知识，这是下一章将要讨论的内容。图2.6-6所示的点集所描述的事件的概率仍可以在合适的区域上对概率密度(pdf)积分求得。下面用实例加以说明。



(a) 事件A的点集



(b) 事件B的点集

图2.6-6 式(2.6-39)不适用的事件A和B的点集

例2.6-6 (非矩形集的概率)给定$f_{XY}(x, y)=\mathrm{e}^{-(x+y)}u(x)u(y)$，计算$P[(X, Y)\in\mathscr{A}]$，其中$\mathscr{A}$为图2.6-7的阴影部分。区域$\mathscr{A}$描述为$\mathscr{A}=\{(x, y): 0\leqslant x\leqslant 1, |y|\leqslant x\}$，那么

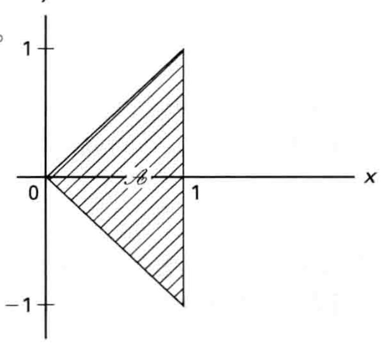
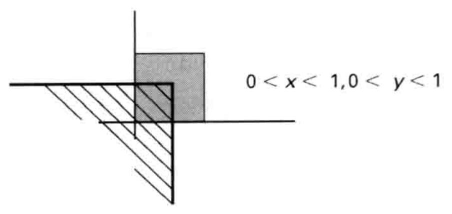


图2.6-7 例2.6-6的区域$\mathscr{A}$

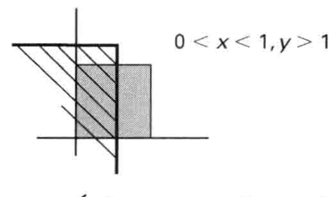
$$
\begin{aligned}
P[(X,Y)\in\mathscr{A}] &= \int_{x=0}^{x=1}\int_{y=-x}^{x}\mathrm{e}^{-(x+y)}u(x)u(y)\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^1\left(\int_{y=-x}^{x}\mathrm{e}^{-y}u(y)\mathrm{d}y\right)\mathrm{e}^{-x}u(x)\mathrm{d}x \\
&= \int_0^1\left(\int_0^{x}\mathrm{e}^{-y}\mathrm{d}y\right)\mathrm{e}^{-x}\mathrm{d}x \\
&= \int_0^1(-\mathrm{e}^{-y}|_0^x)\mathrm{e}^{-x}\mathrm{d}x \\
&= \int_0^1(1-\mathrm{e}^{-x})\mathrm{e}^{-x}\mathrm{d}x \\
&= \int_0^1(\mathrm{e}^{-x}-\mathrm{e}^{-2x})\mathrm{d}x \\
&= \left(-\mathrm{e}^{-x}|_0^1\right)-\left(-\frac{1}{2}\mathrm{e}^{-2x}\right)\Big|_0^1 \\
&= 1-\mathrm{e}^{-1}-\frac{1}{2}+\frac{1}{2}\mathrm{e}^{-2} \\
&= \frac{1}{2}-\mathrm{e}^{-1}+\frac{1}{2}\mathrm{e}^{-2} \\
&\approx 0.1998
\end{aligned}
\tag{2.6-40}
$$

例2.6-7 (CDF计算)令X、Y为两个随机变量，联合概率密度为$f_{XY}(x, y)=1$，$0<x\leqslant 1$，$0<y\leqslant 1$，其他地方为零。概率密度的支撑用灰色表示，事件$(-\infty, x]\times(-\infty, y]$ $(0<x\leqslant 1, 0<y\leqslant 1)$用粗黑线为边界的阴影区域表示。



对于图所示的情况，$F_{XY}(x,\ y)\ =\int_0^y\int_0^x 1\mathrm{d}x'\mathrm{d}y'\ =xy$。

当 $0<x<1,\ y>1$ 时，我们有 $F_{XY}(x,\ y)\ =\int_0^x \mathrm{d}x'\int_0^1 \mathrm{d}y\ =x$，按这种方法继续下去，最终我们得到完整的 CDF 为



$$F_{XY}(x,y)=\begin{cases}0, & x\leqslant 0, y\leqslant 0\\ xy, & 0<x\leqslant 1, 0<y\leqslant 1\\ x, & 0<x\leqslant 1, y>1\\ y, & x>1, 0<y\leqslant 1\\ 1, & x>1, y>1\end{cases}$$

正如例 2.6-6 和例 2.6-7 所表明的特殊情况那样，如果 $f_{XY}(x,\ y)$ 存在，那么任何形如 $\{(X,\ Y)\in\mathscr{A}\}$ 事件的概率都可以由下面的公式计算：

$$P[(X,Y)\in\mathscr{A}]=\iint_{\mathscr{A}} f_{XY}(x,y)\mathrm{d}x\,\mathrm{d}y \tag{2.6-41}$$

式(2.6-41)看上去是合理的，它的真实性还需要证明，一种方法是将任意形状的区域分解成许多小的不相交的矩形区域 $\mathscr{A}_1,\ \mathscr{A}_2,\ \cdots,\ \mathscr{A}_N$，那么事件 $\{X,\ Y\in\mathscr{A}\}$ 被分解为

$$\{(X,Y)\in\mathscr{A}\}=\bigcup_{i=1}^{N}\{(X,Y)\in\mathscr{A}_i\}$$

因此(根据公理 3)

$$P[(X,Y)\in\mathscr{A}]=\sum_{i=1}^{N}P[(X,Y)\in\mathscr{A}_i] \tag{2.6-42}$$

而右边的概率可以用概率分布表示，因而也就可以用概率密度的积分表示式(2.6-39)。取极限，当 N 变得很大时，$\mathscr{A}_i$ 变成无穷小，于是得到式(2.6-41)。

函数 $F_X(x)$ 和 $F_Y(y)$ 如果是从联合分布函数得来，那么也称为边缘分布。因此

$$F_X(x)=F_{XY}(x,\infty)=\int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{XY}(\zeta,y)\mathrm{d}\zeta\mathrm{d}y \tag{2.6-43}$$

$$F_Y(y)=F_{XY}(\infty,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{y} f_{XY}(x,\eta)\mathrm{d}x\mathrm{d}\eta \tag{2.6-44}$$

由于边缘概率密度为

$$f_X(x)=\frac{\mathrm{d}F_X(x)}{\mathrm{d}x} \tag{2.6-45}$$

$$f_Y(y)=\frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} \tag{2.6-46}$$

在式(2.6-43)和式(2.6-44)中分别对 x 和 y 求偏导数，得

$$f_X(x)=\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}y \tag{2.6-47}$$

$$f_Y(y)=\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x \tag{2.6-48}$$

下面将联合 pdf $f_{XY}(x, y)$的一些关键性质总结如下。

联合 pdf 的性质

(i) $f_{XY}(x, y)\geqslant 0$，对所有 x, y。

(ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x\mathrm{d}y = 1$（必然事件）。

(iii) 当$f_{XY}(x, y)$不是一个概率时，它确实可以大于1，我们可以把$f_{XY}(x, y)\mathrm{d}x\mathrm{d}y$ 看成微分概率。有时也写成$f_{XY}(x, y)\mathrm{d}x\mathrm{d}y = P[x < X\leqslant x+\mathrm{d}x, y<Y\leqslant y+\mathrm{d}y]$。

(iv) $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\mathrm{d}y$ 和 $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\mathrm{d}x$。

性质(i)是基于这样的事实得出的：联合 pdf 在整个平面上任意区域上的积分必须是正的。另外，把这个联合 pdf 看成是 CDF 的混合偏导数，性质(i)就可以由联合 CDF 的性质(iv)取极限得出。性质(ii)的得出是基于这样的事实：联合 pdf 在整个平面上的积分给出了随机变量取某些值的概率，随机变量取这些值是必然事件，概率为1。

对于离散随机变量，我们可以得到类似的结果。假定对于所有的 x_i、y_k，联合 PMF 为 $P_{XY}(x_i, y_k)$，边缘 PMF 为

$$P_X(x_i)=\sum_{\text{all } y_k} P_{XY}(x_i,y_k) \tag{2.6-49}$$

$$P_Y(y_k)=\sum_{\text{all } x_i} P_{XY}(x_i,y_k) \tag{2.6-50}$$

例2.6-8 （饭店排队等待时间）某个饭店已经发现新来顾客的等待时间的联合分布以及包含新顾客的总顾客数。令随机变量 W 表示新顾客连续等待时间、离散随机变量 N 表示总的顾客数。

联合分布为

$$F_{W,N}(w,n)=\begin{cases} 0, & n<0 \text{ 或 } w<0 \\ (1-\mathrm{e}^{-w/\mu_0})\frac{n}{10}, & 0\leqslant n<5, w\geqslant 0 \\ (1-\mathrm{e}^{-w/\mu_0})\frac{5}{10}+(1-\mathrm{e}^{-w/\mu_1})\left(\frac{n-5}{10}\right), & 5\leqslant n<10, w\geqslant 0 \\ (1-\mathrm{e}^{-w/\mu_0})\frac{5}{10}+(1-\mathrm{e}^{-w/\mu_1})\left(\frac{5}{10}\right), & 10\leqslant n, w\geqslant 0 \end{cases}$$

其中参数 μ_i 满足 $0<\mu_0<\mu_1$。参数的这种选择意味着顾客越多、等待时间越长。

注意到 W 是一个连续型随机变量，而 N 是一个离散型随机变量，当 $n\geqslant 0$ 和 $w\geqslant 0$ 时，对于某些 n 值，我们将这个联合分布画成 w 的函数，如图 2.6-8 所示。

我们下一步求混合概率质量函数

$$\begin{aligned} f_{W,N}(w,n) &\triangleq \frac{\partial}{\partial w}\nabla_n F_{W,N}(w,n) \\ &= \frac{\partial}{\partial w}\{F_{W,N}(w,n)-F_{W,N}(w,n-1)\} \\ &= \frac{\partial}{\partial w}F_{W,N}(w,n)-\frac{\partial}{\partial w}F_{W,N}(w,n-1) \end{aligned}$$

$f_{W,N}(w, n)$称为 W 与$\{N=n\}$的联合 pdf。通过计算，我们有

$$\nabla_n F_{W,N}(w,n) = F_{W,N}(w,n) - F_{W,N}(w,n-1) = u(w)\begin{cases}(1-\mathrm{e}^{-w/\mu_0})\frac{1}{10}, & 0<n\leqslant 5\\(1-\mathrm{e}^{-w/\mu_1})\frac{1}{10}, & 5<n\leqslant 10\\0, & \text{其他}\end{cases}$$

因此

$$\begin{aligned}f_{W,N}(w,n) &= \frac{\partial}{\partial w}\nabla_n F_{W,N}(w,n)\\&= u(w)\begin{cases}\frac{1}{10}\frac{1}{\mu_0}\mathrm{e}^{-w/\mu_0}, & 0<n\leqslant 5\\\frac{1}{10}\frac{1}{\mu_1}\mathrm{e}^{-w/\mu_1}, & 5<n\leqslant 10\\0, & \text{其他}\end{cases}\end{aligned}$$

由联合 pdf，我们看到一个简单的观点，在 n 小于或等于 5 个顾客时，较小的平均等待时间 μ_0 控制着随机变量 W，而当 n 大于 5 个顾客时，较大的平均等待时间 μ_1 控制着随机变量 W。在更为详细的模型中，平均等待时间将随 n 的增加而上升。

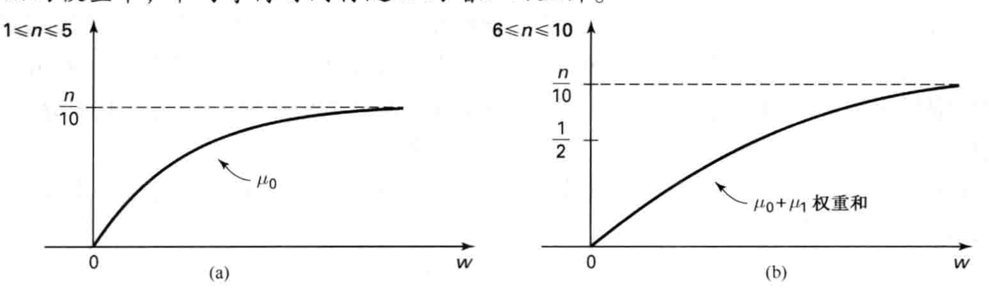


图 2.6-8 例 2.6-8 的 CDF。(a)范围在 1 ~5 的顾客数；(b) 范围在 6 ~10 的顾客数

独立随机变量 对于两个随机变量 X 和 Y，如果事件$\{X\leqslant x\}$和$\{Y\leqslant y\}$对 x 和 y 的每一个组合都是独立的，则称 X 和 Y 是独立的。在 1.5 节中，对于事件 A 和 B，如果 $P[AB]=P[A]P[B]$，则称 A 和 B 是独立的。取 $AB\triangleq\{X\leqslant x\}\cap\{Y\leqslant y\}$，其中 $A\triangleq\{X\leqslant x\}$，$B=\{Y\leqslant y\}$，回想一下，$F_X(x)\triangleq P[X\leqslant x]$，对 $F_Y(y)$ 也类似。对于每一个 x 和 y，当且仅当 X 与 Y 独立时，可得出

$$F_{XY}(x,y) = F_X(x)F_Y(y) \tag{2.6-51}$$

另外

$$\begin{aligned}f_{XY}(x,y) &= \frac{\partial^2 F_{XY}(x,y)}{\partial x\,\partial y}\\&= \frac{\partial F_X(x)}{\partial x}\cdot\frac{\partial F_Y(y)}{\partial y}\end{aligned} \tag{2.6-52}$$

$$= f_X(x)f_Y(y) \tag{2.6-53}$$

由条件概率的定义，对于独立的 X 和 Y，有

$$\begin{aligned}F_X(x|Y\leqslant y) &= \frac{F_{XY}(x,y)}{F_Y(y)}\\&= F_X(x)\end{aligned} \tag{2.6-54}$$

对 $F_Y(y|\ X\leqslant x)$ 也是类似的。对于独立的事件，由这些结果(通过求导)可以得出条件 pdf 等于边缘 pdf，即

$$f_X(x|Y\leqslant y) = f_X(x) \tag{2.6-55}$$

$$f_Y(y|X\leqslant x) = f_Y(y) \tag{2.6-56}$$

由式(2.6-39)很容易证明：如果X和Y是两个独立的随机变量，那么事件$\{x_1 < X \leqslant x_2\}$和$\{y_1 < Y \leqslant y_2\}$也是独立的，即

$$P[x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2] = P[x_1 < X \leqslant x_2]P[y_1 < Y \leqslant y_2] \tag{2.6-57}$$

如果$F_{XY}(x, y) = F_X(x)F_Y(y)$，利用式(2.6-39)，有

$$P[x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2]$$
$$= F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1) \tag{2.6-58}$$
$$= F_X(x_2)F_Y(y_2) - F_X(x_2)F_Y(y_1) - F_X(x_1)F_Y(y_2) + F_X(x_1)F_Y(y_1) \tag{2.6-59}$$
$$= (F_X(x_2) - F_X(x_1))(F_Y(y_2) - F_Y(y_1)) \tag{2.6-60}$$
$$= P[x_1 < X \leqslant x_2]P[y_1 < Y \leqslant y_2] \tag{2.6-61}$$

例2.6-9 假定随机实验为投掷一枚均匀骰子一次，实验的样本空间为$\Omega = \{1, 2, 3, 4, 5, 6\}$。我们定义两个随机变量如下：

$$X(\zeta) \triangleq \begin{cases} 1+\zeta & \text{实验结果为 1 或 3} \\ 0 & \text{其他实验结果} \end{cases}$$

$$Y(\zeta) \triangleq \begin{cases} 1-\zeta & \text{实验结果为 1, 2 或 3} \\ 0 & \text{其他实验结果} \end{cases}$$

(a) 计算单个的和联合的 PMF。

(b) 计算联合 CDF $F_{XY}(1, 1)$、$F_{XY}(3, -0.5)$、$F_{XY}(5, -1.5)$。

(c) 随机变量X和Y是独立的吗？

解 由于骰子假定是均匀的，每一面向上的概率为1/6。

(a) 所以单一事件$\{\zeta\}$是等可能的，概率为$P[\{\zeta\}] = 1/6$，因此

$$X(1) = 2,\ X(3) = 4，\text{对其他实验结果，有}$$
$$X(2) = X(4) = X(5) = X(6) = 0$$

于是，PMF 的P_X为$P_X(0) = 4/6$，$P_X(2) = 1/6$，$P_X(4) = 1/6$，对所有其他k，$P_X(k) = 0$。

类似地，由$Y(\zeta)$的定义，有

$$Y(1) = Y(4) = Y(5) = Y(6) = 0$$
$$Y(2) = -1 \text{ 和 } Y(3) = -2$$

由此可得 PMF 的值$P_Y(0) = 4/6$，$P_Y(-1) = 1/6$，$P_Y(-2) = 1/6$，对所有其他k，$P_Y(k) = 0$。

接下来直接根据定义计算联合 PMF $P_{XY}(i, j)$，即$P_{XY}(i, j) = P[\text{所有}\zeta: X(\zeta) = i, Y(\zeta) = j]$。如果我们回想一下联合概率是$\Omega$的子集的交集，那么这项计算是很容易的，例如考察骰子的面为2、4、5或6，将其写成子集$\{2, 4, 5\}$，于是，$P_{XY}(0, 0) = P[\text{所有}\zeta: X(\zeta) = 0, Y(\zeta) = 0] = P[\{2, 4, 5, 6\} \cap \{1, 4, 5, 6\}] = P[\{4, 5, 6\}] = 1/2$。

类似地，我们计算

$$P_{XY}(2,0) = P[\{1\} \cap \{1,4,5,6\}] = P[\{1\}] = 1/6$$
$$P_{XY}(4,0) = P[\{3\} \cap \{1,4,5,6\}] = P[\phi] = 0$$
$$P_{XY}(0,-1) = P[\{2,4,5,6\} \cap \{2\}] = P[\{2\}] = 1/6$$
$$P_{XY}(2,-1) = P[\{1\} \cap \{2\}] = P[\phi] = 0$$
$$P_{XY}(4,-1) = P[\{3\} \cap \{2\}] = P[\phi] = 0$$
$$P_{XY}(0,-2) = P[\{2,4,5,6\} \cap \{3\}] = P[\phi] = 0$$
$$P_{XY}(2,-2) = P[\{1\} \cap \{3\}] = P[\phi] = 0$$
$$P_{XY}(4,-2) = P[\{3\} \cap \{3\}] = P[\{3\}] = 1/6$$

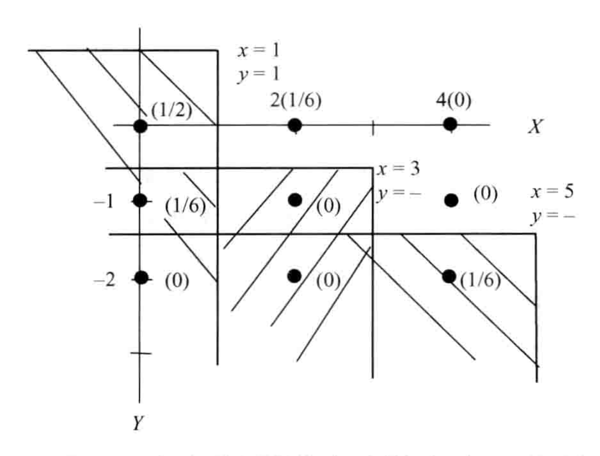
(b) 为了计算联合 CDF, 把这些点和关联的概率画出来是有帮助的, 这些概率在括号里标注出来, 从图可以看出

$$F_{XY}(1,1)=P_{XY}(0,0)+P_{XY}(0,1)+P_{XY}(0,2)=2/3$$

类似地, $F_{XY}(3,\ -0.5)=P_{XY}(0,\ -1)+P_{XY}(2,\ -1)+P_{XY}(0,\ -2)+P_{XY}(2,\ -2)=\frac{1}{6}$

和 $F_{XY}(5,\ -1.5)=P_{XY}(0,\ -2)+P_{XY}(2,\ -2)+P_{XY}(4,\ -2)=\frac{1}{6}$。

(c) 要考察非独立性, 只要找到一点, 其中的 pdf(或 CDF) 不能分解就足够了。考虑 $P_{XY}(2,0)=1/6$, 而 $P_X(2)P_Y(0)=1/6\times 4/6=1/9$, 所以, 随机变量 X 和 Y 不是独立的。



与例 2.6-9 有关的概率, 括号中的数是到一点的概率, 如 $P_{XY}(0,0)=1/2$

例 2.6-10 (独立高斯随机变量的联合 pdf)

$$\begin{aligned} f_{XY}(x,y) &= \frac{1}{2\pi\sigma^2}\mathrm{e}^{-(1/2\sigma^2)(x^2+y^2)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{1}{2}(x^2/\sigma^2)}\frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{1}{2}(y^2/\sigma^2)} \end{aligned} \tag{2.6-62}$$

因此 X 和 Y 是独立的随机变量。

例 2.6-11 (独立高斯随机变量的计算) 两个随机变量的联合 pdf 为 $f_{XY}(x,y)=[2\pi]^{-1}\exp\left[-\frac{1}{2}(x^2+y^2)\right]$, $-\infty<x,\ y<\infty$, 计算 X 和 Y 被限制在下列区域里的概率 (a) 2×2 区域; (b) 单位圆。

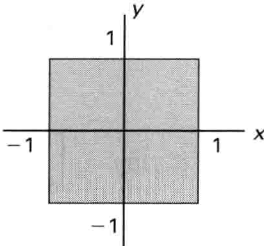
解

(a) 令 $\Re_1$ 表示正方形平面, 那么

$$P[\zeta:(X,Y)\in\Re_1]=\iint_{\Re_1}f_{XY}(x,y)\mathrm{d}x\mathrm{d}y \tag{2.6-63}$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-1}^{1}\exp\left[-\frac{1}{2}x^2\right]\mathrm{d}x\times\frac{1}{\sqrt{2\pi}}\int_{-1}^{1}\exp\left[-\frac{1}{2}y^2\right]\mathrm{d}y \tag{2.6-64}$$

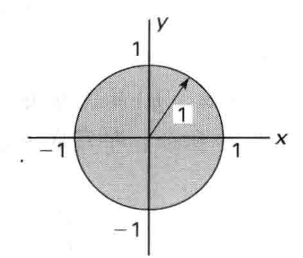
$$= 2\text{erf}(1) \times 2\text{erf}(1) = 0.465 \tag{2.6-65}$$



(b)令 $\Re_2$ 表示单位圆的平面，那么

$$P[\zeta: (X,Y) \in \Re_2] = \iint_{\Re_2} f_{XY}(x,y)\mathrm{d}x\mathrm{d}y \tag{2.6-66}$$

$$= \iint_{\Re_2} [2\pi]^{-1} \exp\left[-\frac{1}{2}(x^2+y^2)\right] \mathrm{d}x\mathrm{d}y \tag{2.6-67}$$



代入 $r \triangleq \sqrt{x^2+y^2}$ 和 $\tan\theta \triangleq y/x$，无穷小的区域 $\mathrm{d}x\mathrm{d}y \to r\mathrm{d}r\mathrm{d}\theta$ 得

$$P[\zeta: (X,Y) \in \Re_2] = \frac{1}{2\pi} \iint_{\Re_2} \exp\left(-\frac{1}{2}r^2\right) r\,\mathrm{d}r\,\mathrm{d}\theta \tag{2.6-68}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left[\int_0^1 r\exp\left(-\frac{1}{2}r^2\right)\mathrm{d}r\right]\mathrm{d}\theta \tag{2.6-69}$$

$$= \int_0^1 r\exp\left(-\frac{1}{2}r^2\right)\mathrm{d}r \tag{2.6-70}$$

$$= \int_0^{1/2} \mathrm{e}^{-z}\mathrm{d}z, \qquad z \triangleq \frac{1}{2}r^2 \tag{2.6-71}$$

$$\doteq 0.393 \tag{2.6-72}$$

含有非独立随机变量的联合概率密度　留给读者思考所有联合 CDF 或 pdf 的分解，下面考虑一种含有非独立随机变量的情况。

例 2.6-12　(计算联合 CDF)考虑一种简单但不可分解的联合 pdf

$$f_{XY}(x,y) = A(x+y) \qquad 0 < x \leqslant 1, \qquad 0 < y \leqslant 1 \tag{2.6-73}$$

$$= 0, \qquad \text{其他} \tag{2.6-74}$$

回答下列问题。

(i) A 是多少？我们知道

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x\,\mathrm{d}y = 1$$

因此

$$A\int_0^1 \mathrm{d}y\int_0^1 x\,\mathrm{d}x + A\int_0^1 \mathrm{d}x\int_0^1 y\,\mathrm{d}y = 1 \Rightarrow A = 1$$

(ii) 边缘 pdf 是多少?

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}y = \int_0^1 (x+y)\mathrm{d}y = (xy + y^2/2)\Big|_0^1$$
$$= \begin{cases} x+\frac{1}{2}, & 0 < x \leqslant 1 \\ 0, & \text{其他} \end{cases}$$

类似地

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\mathrm{d}x$$
$$= \begin{cases} y+\frac{1}{2}, & 0 < y \leqslant 1 \\ 0, & \text{其他} \end{cases}$$

(iii) $F_{XY}(x, y)$是多少? $F_{XY}(x, y) \triangleq P[X\leqslant x, Y\leqslant y]$,所以,我们必须在顶点为$(x, y)$, $(x, -\infty)$, $(-\infty, -\infty)$和$(-\infty, y)$的无限矩形上积分。然而,只是在这个矩形与$f_{XY}(x, y)\neq 0$的区域重叠部分积分,即写成supp(f)的pdf的支撑区域将是对积分有贡献的

$$F_{XY}(x,y) = \int_{-\infty}^{x} \mathrm{d}x' \int_{-\infty}^{y} \mathrm{d}y' f_{XY}(x',y')$$

(a) $x\geqslant 1$, $y\geqslant 1$[参见图2.6-9(a)]

$$F_{XY}(x,y) = \int_0^1 \int_0^1 f_{XY}(x',y')\mathrm{d}x'\,\mathrm{d}y' = 1$$

(b) $0 < x\leqslant 1$, $y\geqslant 1$[参见图2.6-9(b)]

$$F_{XY}(x,y) = \int_{y'=0}^{1} \mathrm{d}y' \left(\int_{x'=0}^{x} \mathrm{d}x'(x'+y')\right)$$
$$= \int_{y'=0}^{1} \mathrm{d}y' \left(\int_{x'=0}^{x} x'\,\mathrm{d}x\right) + \int_{y'=0}^{1} \mathrm{d}y'\, y' \left(\int_{x'=0}^{x} \mathrm{d}x'\right)$$
$$= \frac{x}{2}(x+1)$$

(c) $0 < y\leqslant 1$, $x\geqslant 1$ [参见图2.6-9(c)]

$$F_{XY}(x,y) = \int_{y'=0}^{y} \int_{x'=0}^{1} (x'+y')\mathrm{d}x'\,\mathrm{d}y' = \frac{y}{2}(y+1)$$

(d) $0 < x\leqslant 1$, $0 < y\leqslant 1$ [参见图2.6-9(d)]

$$F_{XY}(x,y) = \int_{y'=0}^{y} \int_{x'=0}^{x} (x'+y')\mathrm{d}x'\,\mathrm{d}y' = \frac{yx}{2}(x+y)$$

(e) $x < 0$, y任意;或者$y < 0$, x任意 [参见图2.6-9(e)]

$$F_{XY}(x,y) = 0$$

(f) 计算$P[X+Y\leqslant 1]$。点集是由直线$x+y=1$或$y=1-x$分开的半空间。然而,只有这个半空间与$f_{XY}(x, y)\neq 0$的那个区域相交的部分才对积分有贡献[参见图2.6-9(f)],因此

$$P[X+Y\leqslant 1]=\iint_{x'+y'\leqslant 1} f_{XY}(x',y')\mathrm{d}x'\,\mathrm{d}y'$$

$$\begin{aligned}P[X+Y\leqslant 1]&=\int_{x'=0}^{1}\int_{y'=0}^{1-x'}(x'+y')\mathrm{d}y'\,\mathrm{d}x'\\&=\int_{x'=0}^{1}x'(1-x')\mathrm{d}x'+\int_{x'=0}^{1}\frac{(1-x')^2}{2}\mathrm{d}x'\\&=\frac{1}{3}\end{aligned}$$

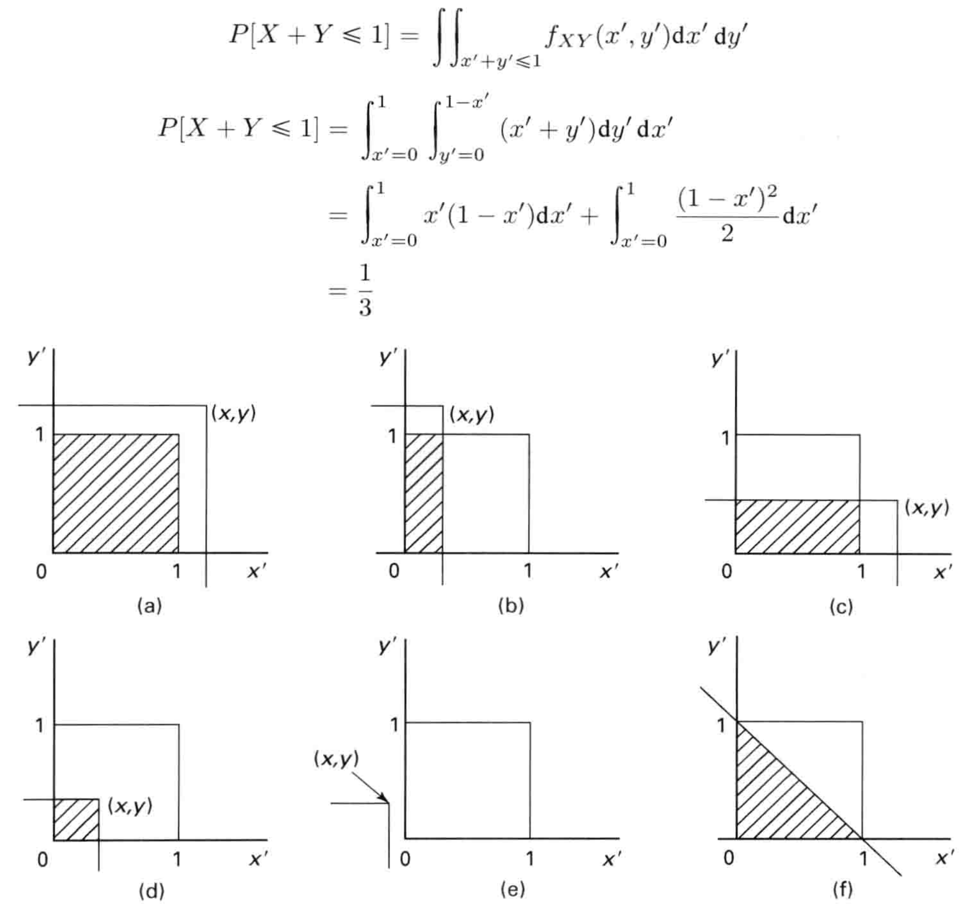


图 2.6-9 (a)到(e)的阴影区域是 $\mathrm{supp}(f_{XY})$ 与事件$\{-\infty<X\leqslant x,\ -\infty<Y\leqslant y\}$有关的点集的交集，在(f)中，阴影部分是 $\mathrm{supp}(f_{XY})$ 与$\{X+Y\leqslant 1\}$的交集

在前一个例子中，我们讨论了不可分解的 pdf，联合 pdf 不可分解的另一个例子是

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}}\exp\left(\frac{-1}{2\sigma^2(1-\rho^2)}(x^2+y^2-2\rho xy)\right) \qquad (2.6\text{-}75)$$

当 $\rho\neq 0$ 时，X 和 Y 不是独立的。

在式(2.6-75)中，当 $\rho=0$ 时，$f_{XY}(x,\ y)$分解为$f_X(x)f_Y(y)$，X 和 Y 变成了独立的随机变量。$f_{XY}(x,\ y)$的图形如图 2.6-10 所示，图中$\sigma=1$。

在第 4 章将会看到，式(2.6-75)是两个随机变量的联合高斯概率密度的特例，在讨论了参数 ρ 的意义以后，再对这一重要的 pdf 做进一步的讨论，这一工作将在第 4 章进行。在第 5 章会看到式(2.6-75)及其推广能够用矩阵形式紧凑地表示。

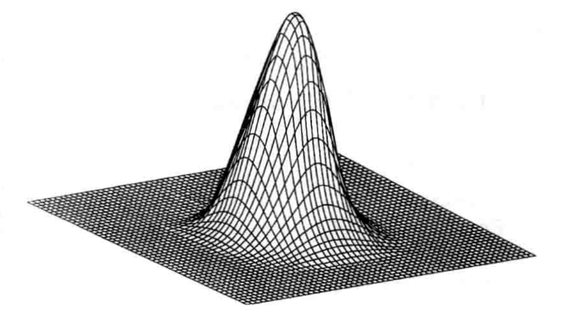


图 2.6-10 联合高斯概率密度的图形

$$f_{XY}(x,y)=(2\pi)^{-1}\exp\left[-\frac{1}{2}(x^2+y^2)\right]$$

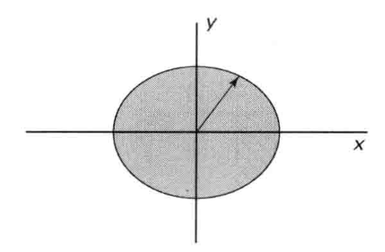
例 2.6-13 (非独立高斯随机变量的计算)再次考虑例 2.6-11(b)的情况，但联合 pdf 为

$$f_{XY}(x,y)=[2\pi\sqrt{1-\rho^2}]^{-1}\exp\left[-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right]$$

和前面一样，令 $\Re_2$ 表示单位圆的平面，那么

$$P[\zeta:(X,Y)\in\Re_2]=\iint_{\Re_2}f_{XY}(x,y)\mathrm{d}x\mathrm{d}y \tag{2.6-76}$$

$$=\iint_{\Re_2}[2\pi\sqrt{1-\rho^2}]^{-1}\exp\left(-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right)\mathrm{d}x\,\mathrm{d}y \tag{2.6-77}$$



用极坐标替换，$r\triangleq\sqrt{x^2+y^2}$，$\tan\theta\triangleq y/x$，得

$$P[\zeta:(X,Y)\in\Re_2]=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_0^{2\pi}\int_0^1 r\exp\left(-\frac{1}{2(1-\rho^2)}(r^2-2\rho r^2\cos\theta\sin\theta)\right)\mathrm{d}r\mathrm{d}\theta$$

$$=\frac{K}{\pi}\int_0^{2\pi}\int_0^1 r\exp\left(-r^2\left[2K^2(1-\rho\sin 2\theta)\right]\right)\mathrm{d}r\mathrm{d}\theta$$

$$K\triangleq\frac{1}{2\sqrt{1-\rho^2}}\quad 和\quad \sin 2\theta=2\sin\theta\cos\theta$$

$$=\frac{K}{2\pi}\int_0^{2\pi}\left[\int_0^1\exp-\left(\left[2K^2(1-\rho\sin 2\theta)\right]z\right)\,dz\right]\mathrm{d}\theta\quad 且\quad z\triangleq r^2$$

$$=\frac{K}{2\pi}\int_0^{2\pi}\frac{1}{2K^2(1-\rho\sin 2\theta)}(1-\exp-\left[2K^2(1-\rho\sin 2\theta)\right]\mathrm{d}\theta$$

$$=\frac{1}{4\pi K}\int_0^{2\pi}\left(\frac{1-\exp[-2K^2(1-\rho\sin 2\theta)]}{1-\rho\sin 2\theta}\right)\mathrm{d}\theta$$

当 $\rho=0$ 时，我们得到概率为 0.393，与例 2.6-10 的结果相同。然而，当 $\rho\neq 0$ 时，积分不能用闭合形式表示，这个概率必须进行数值计算，下面提供了计算 $P(\zeta:(X,Y)\in\Re_2)$ 的 MATLAB 程序，结果显示在图 2.6-11 中。

MATLAB. M file for computing. $P[\zeta:(X,Y)\in\Re_2]$

```
function[Pr]=corrprob
p=[0:100]/100.;
q=p*2*pi;
Pr=zeros(1,100);

K=.5./sqrt(1-p.^2);

for i=1:100
   f=(1-exp(-2*K(i)^2*(1-p(i)*sin(2*q)))./(1-p(i)*sin(2*q));
   Pr(i)=sum(f)/(4*pi)/K(i)*(2*pi/100);
end
```

```
plot(p(1:100),Pr)
title('Probability that two correlated Gaussian RVs take values in the
  unit circle')
xlabel('Correlation coefficient rho')
ylabel('Probability that X,Y) lie in the unit circle')
```

在4.3节，我们证明了当$\rho\to1$时，$f_{XY}(x,y)\to\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\delta(y-x)$，因此

$$P[\zeta:(X,Y)\in\Re_2]=\iint_{\Re_2}\frac{1}{\sqrt{2\pi}} \tag{2.6-78}$$

$$\mathrm{e}^{-0.5x^2}\delta(x-y)\mathrm{d}x\mathrm{d}y=\int_{-0.707}^{0.707}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-0.5x^2}\mathrm{d}x \tag{2.6-79}$$

$$=0.52 \tag{2.6-80}$$

这一结果在图2.6-11中也能看到。

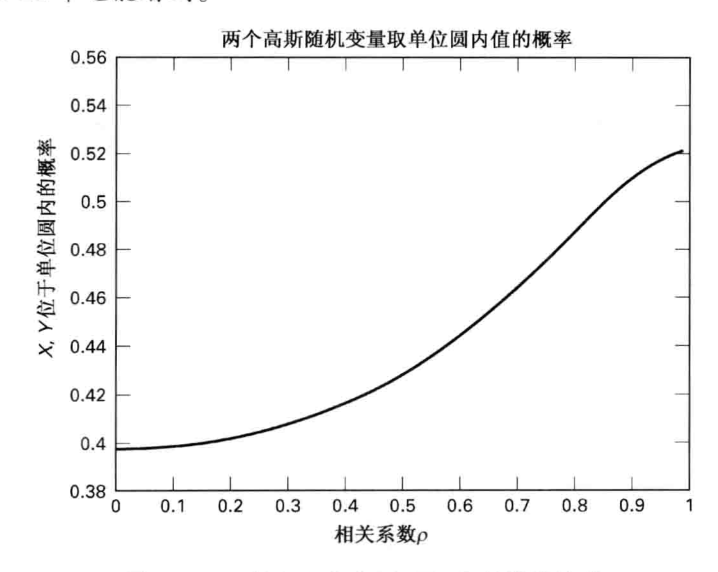


图2.6-11 例2.6-13中MATLAB的计算结果

条件密度 我们现在推导两个随机变量的条件密度的一个有用公式。这个公式是基于式(1.6-2)给出的条件概率，由式(2.6-39)，我们有

$$\begin{aligned}&P[x<X\leqslant x+\Delta x,y<Y\leqslant y+\Delta y]\\&=F_{XY}(x+\Delta x,y+\Delta y)-F_{XY}(x,y+\Delta y)-F_{XY}(x+\Delta x,y)+F_{XY}(x,y)\end{aligned} \tag{2.6-81}$$

两边除以$\Delta x\Delta y$并取极限，这时右边可以看成$F_{XY}(x,y)$对x和y的偏导数，可以写成

$$\lim_{\substack{\Delta x\to0\\\Delta y\to0}}\frac{P[x<X\leqslant x+\Delta x,y<Y\leqslant y+\Delta y]}{\Delta x\,\Delta y}=\frac{\partial^2F_{XY}}{\partial x\,\partial y}\triangleq f_{XY}(x,y)$$

因此，对于小的Δx、Δy

$$P[x<X\leqslant x+\Delta x,y<Y\leqslant y+\Delta y]\approx f_{XY}(x,y)\Delta x\,\Delta y \tag{2.6-82}$$

这是式(2.4-6)在二维时的等价形式。考虑

$$P[y < Y \leqslant y + \Delta y | x < X \leqslant x + \Delta x] = \frac{P[x < X \leqslant x + \Delta x, y < Y \leqslant y + \Delta y]}{P[x < X \leqslant x + \Delta x]} \approx \frac{f_{XY}(x,y)\Delta x \, \Delta y}{f_X(x)\Delta x} \tag{2.6-83}$$

但左边的量是

$$F_{Y|B}(y + \Delta y | x < X \leqslant x + \Delta x) - F_{Y|B}(y | x < X \leqslant x + \Delta x)$$

其中 $B \triangleq \{x < X \leqslant x + \Delta x\}$，因此，由式(2.6-3)，有

$$\begin{aligned} &\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{F_{Y|B}(y + \Delta y | x < X \leqslant x + \Delta x) - F_{Y|B}(y | x < X \leqslant x + \Delta x)}{\Delta y} \\ &= \frac{f_{XY}(x,y)}{f_X(x)} \\ &= \frac{\partial F_{X|Y}(y|X = x)}{\partial y} \\ &= f_{Y|X}(y|x) \end{aligned} \tag{2.6-84}$$

$f_{Y|X}(y|x)$的表示形式提醒我们这是事件$\{X = x\}$给定的条件下 Y 的条件 pdf。于是，可得到一个重要的公式

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}, \qquad f_X(x) \neq 0 \tag{2.6-85}$$

如果在式(2.6-48)中使用式(2.6-85)，则可以得到一个有用的公式

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) \mathrm{d}x \tag{2.6-86}$$

又

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}, \qquad f_Y(y) \neq 0 \tag{2.6-87}$$

$f_{X|Y}(x|y)$称为事件$\{Y = y\}$给定的条件下 X 的条件 pdf。由式(2.6-85)和式(2.6-86)可见

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} \tag{2.6-88}$$

下面用一个例子加以说明。

例 2.6-14 (激光相干性)假定我们观测到一束由激光器发射的光场 $U(t)$，激光是时间上相干的，也就是对于任意两个时刻 t_1 和 t_2，如果 $t_2 - t_1$ 不是很大，光就不是统计独立的[2-5]，令 $X \triangleq U(t_1)$，$Y \triangleq U(t_2)$ 且 $t_2 > t_1$。假定 X 和 Y 可以看成是联合高斯的①，联合 pdf 如式(2.6-75)所示，但 $\sigma^2 = 1$。当 $\rho \neq 0$ 时，X 和 Y 不是独立的，利用式(2.6-47)和式(2.6-48)的定义，可以证明边缘概率密度 $f_X(x)$ 和 $f_Y(y)$ 分别都是高斯的，我们将在第 4 章中证明这一点。由于均值都为 0、方差都为 1，于是可得边缘概率密度

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2} \qquad 和 \qquad f_Y(y) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}y^2}$$

---

① 由于光子特性，光常用泊松分布来建模，正如在第 1 章中看到的那样，对于大的光子计数器，高斯分布可以很好地近似泊松分布。当然，光密度不能为负，但如果均值与标准差相比很大时($\mu \gg \sigma$)，高斯概率密度在那里是非常小的。

都是以零为中心的。现在假定在 t_1 时刻测得的光为 X，即 $X=x$，在这些新知识情况下，Y 的 pdf 仍是以零为中心吗？即 Y 的平均值①仍是零吗？

解 我们希望计算 $f_{Y|X}(y|x)$。应用式(2.6-85)

$$f_{Y|X}(y|x)=\frac{f_{XY}(x,y)}{f_X(x)}$$

得

$$f_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left\{\left[-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right]+\frac{1}{2}x^2\right\}$$

将上式整理后，得

$$f_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left\{-\left[\frac{x^2+y^2-2\rho xy-x^2(1-\rho^2)}{2(1-\rho^2)}\right]\right\}$$

将指数里的二次项组合成完美的平方项，得

$$f_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right)$$

因此，当 $X=x$ 时，Y 的 pdf 是以 $y=\rho x$ 为中心的，很显然不是零。如果 $\rho x>0$，Y 有更多的可能性取正值，如果 $\rho x<0$，Y 有更多的可能性取负值。这与 X 没有观察到的情况形成鲜明的对照，这时的 Y 值更多的可能性是零。

故障率的估计是条件事件和条件概率的主要应用，这在下一节中讨论。

2.7 故障率

在现代工业化社会，设备更新、人寿保险的发售等重要活动，都有目标故障率的仔细记录，这些目标是机器或人。例如，考虑人寿保险的成本：给一个 25 岁的妇女确定的人寿保险价格与一个 75 岁的老人处于同一水平是不划算的。对于老人的"故障"概率(即死亡)要比年轻妇女大得多。因此，合理的定价策略就要求保险公司以更高的价格为老人保险，应该高多少呢？这可以由精算表确定，精算表是在多种条件下平均寿命的估计。一个重要的条件是"你一直活到(某个年龄)"。换句话说，在给定你已经活到 85 岁的条件下，你将活到 86 岁的概率要远大于你还是婴儿的条件下你将活到 86 岁的概率。

令 X 表示故障的时间，由贝叶斯定理，在目标已经存活到 t 的情况下，故障将在间隔 $[t, t+\mathrm{d}t]$ 上发生的概率可以写成

$$P[t<X\leqslant t+\mathrm{d}t|X>t]=\frac{P[t<X\leqslant t+\mathrm{d}t, X>t]}{P[X>t]} \tag{2.7-1}$$

由于事件 $\{X>t\}$ 包含了事件 $\{t<X\leqslant t+\mathrm{d}t\}$，可得出 $P[t<X\leqslant t+\mathrm{d}t, X>t]=P[t<X\leqslant t+\mathrm{d}t]$。因此

$$P[t<X\leqslant t+\mathrm{d}t|X>t]=\frac{P[t<X\leqslant t+\mathrm{d}t]}{P[X>t]} \tag{2.7-2}$$

注意到 $P[t<X\leqslant t+\mathrm{d}t]=F_X(t+\mathrm{d}t)-F_X(t)$，于是

① 完整的概念将在第 4 章中给出。

$$P[t < X \leqslant t+\mathrm{d}t|X>t] = \frac{F_X(t+\mathrm{d}t)-F_X(t)}{1-F_X(t)} \tag{2.7-3}$$

CDF $F_X(t+\mathrm{d}t)$在点 t 处用泰勒级数展开(假定 F_X 是可导的)，得

$$F_X(t+\mathrm{d}t) = F_X(t)+f_X(t)\,\mathrm{d}t$$

将这一结果应用到式(2.7-3)中，得

$$\begin{aligned} P[t < X \leqslant t+\mathrm{d}t|X>t] &= \frac{f_X(t)\mathrm{d}t}{1-F_X(t)} \\ &\triangleq \alpha(t)\,\mathrm{d}t \end{aligned} \tag{2.7-4}$$

其中

$$\alpha(t) \triangleq \frac{f_X(t)}{1-F_X(t)} \tag{2.7-5}$$

目标 $\alpha(t)$称为条件故障率，尽管它还有其他的名字，如冒险率、死亡率、密度率、瞬时故障率或简单地称为故障率等。如果 t 时刻的条件故障率大，那么存活到 t 的目标在下一个 Δt 秒要比低条件故障率的目标有更大的故障概率。包括人在内的许多目标都有一种随时间变化的故障率。在目标的早期生活期间，由于固有的或先天的缺陷，故障率可能比较高。在这个早期过去以后，目标可能享有近似为常数故障率的有用生活期。最后，随着年龄的增加和部件的损耗，故障率就会显著增加而导致快速而又冷酷地发生故障或死亡。

注意到 $F_X(t+\mathrm{d}t)-F_X(t)=F'_X(t)\,\mathrm{d}t=\mathrm{d}F_X$，随机变量 X 的 pdf 可以按式(2.7-3)计算，于是有

$$\frac{\mathrm{d}F_X}{1-F_X} = \alpha(t)\,\mathrm{d}t \tag{2.7-6}$$

通过积分可以求解上式。首先回顾一下积分

$$\int_{y_0}^{y_1} \frac{dy}{1-y} = -\int_{1-y_0}^{1-y_1} \frac{\mathrm{d}y}{y} = \int_{1-y_1}^{1-y_0} \frac{\mathrm{d}y}{y} = \ln\frac{1-y_0}{1-y_1}$$

其次，利用下列事实:

(i) $F_X(0)=0$，因为我们假定目标在 $t=0$(目标开机的时间)开始工作。

(ii) $F_X(\infty)=1$，因为我们假定目标最终是要坏掉。那么

$$\int_{F_X(0)}^{F_X(t)} \frac{\mathrm{d}F_X}{1-F_X} = -\ln[1-F_X(t)] = \int_0^t \alpha(t')\mathrm{d}t'$$

最终可得

$$F_X(t) = 1-\exp\left[-\int_0^t \alpha(t')\mathrm{d}t'\right] \tag{2.7-7}$$

由于 $F_X(\infty)=1$，所以肯定有

$$\int_0^{\infty} \alpha(t')\,\mathrm{d}t' = \infty \tag{2.7-8}$$

式(2.7-7)是故障时间 X 的 CDF，对式(2.7-7)求导，我们得到 pdf

$$f_X(t) = \alpha(t)\exp\left[-\int_0^t \alpha(t')\mathrm{d}t'\right] \tag{2.7-9}$$

条件故障率 $\alpha(t)$的不同模型导出不同的 pdf。

例2.7-1 （指数型的条件故障率）假定 X 服从指数型的概率分布，即 $F_X(t)=(1-\mathrm{e}^{-\lambda t})u(t)$，则

$$\alpha(t)=\frac{f_X(t)}{1-F_X(t)}=\frac{\lambda\mathrm{e}^{-\lambda t}}{\mathrm{e}^{-\lambda t}}=\lambda$$

因此，条件故障率是一个常数。相反，如果 $\alpha(t)$ 是常数，故障时间就服从指数概率分布。

要注意的重要点是条件故障率并不是pdf[参见式(2.7-8)]。给定 $\{X\geqslant t\}$，X 的条件密度可对条件分布求导得到，例如

$$\begin{aligned}F_X(x|X>t)&\triangleq P[X\leqslant x|X>t]\\&=\frac{P[X\leqslant x,X>t]}{P[X>t]}\end{aligned}\tag{2.7-10}$$

很显然，如果 $t>x$，那么事件 $\{X\leqslant x,\ X>t\}$ 就是一个空集，如果 $t<x$，那么 $\{X\leqslant x,\ X>t\}=\{t<X\leqslant x\}$，于是

$$F_X(x|X>t)=\begin{cases}0, & t>x\\ \dfrac{F_X(x)-F_X(t)}{1-F_X(t)}, & x\geqslant t\end{cases}\tag{2.7-11}$$

因此

$$f_X(x|X\geqslant t)=\begin{cases}0, & t>x\\ \dfrac{f_X(x)}{1-F_X(t)}, & x\geqslant t\end{cases}\tag{2.7-12}$$

$\alpha(t)$ 与 $f_X(x|X\geqslant t)$ 之间的联系是通过比较式(2.7-12)与式(2.7-5)得到的，即

$$f_X(t|X\geqslant t)=\alpha(t)\tag{2.7-13}$$

例2.7-2 （Itsibitsi 的故障）奥斯卡是一名大学生，他有一辆使用了9年的 Itsibitsi 汽车，这是一辆由于可靠性高而著名的进口汽车。假定在正常行驶10 000英里/年的情况下，基于该领域数据的条件故障率为 $\alpha(t)=0.06tu(t)$，为了庆祝他毕业，奥斯卡开始了30天横跨国家的汽车旅行。奥斯卡的 Itsibitsi 汽车在这次旅行中第一次出故障的概率是多少？

解 首先计算pdf $f_X(t)$

$$f_X(t)=0.06t\mathrm{e}^{-\int_0^t 0.06t'\,\mathrm{d}t'}u(t)\tag{2.7-14}$$

$$=0.06t\mathrm{e}^{-0.03t^2}u(t)\tag{2.7-15}$$

其次，将30天转换成0.0824年。最后，我们注意到

$$P[9.0<X\leqslant 9.0824|X>9]=\frac{P[9.0<X\leqslant 9.0824]}{1-F_X(9)}$$

其中，我们利用了贝叶斯公式以及事件 $\{9<X\leqslant 9.0824\}\cap\{X>9\}=\{9<X\leqslant 9.0824\}$，由于

$$P[9.0<X\leqslant 9.0824]=0.06\int_{9.0}^{9.0824}t\mathrm{e}^{-0.03t^2}\mathrm{d}t\tag{2.7-16}$$

$$=\frac{0.06}{2}\int_{(9.0)^2}^{(9.0824)^2}\mathrm{e}^{-0.03z}\mathrm{d}z \quad 并且 \quad z\triangleq t^2\tag{2.7-17}$$

$$=\frac{0.06}{2}\frac{1}{0.03}\left(\mathrm{e}^{-0.03(9.0)^2}-\mathrm{e}^{-0.03(9.0824)^2}\right)\tag{2.7-18}$$

$$= \left(e^{-0.03(9.0)^2} - e^{-0.03(9.0824)^2}\right) \tag{2.7-19}$$

$$\approx 0.0038 \tag{2.7-20}$$

和

$$1 - F_X(9) = 0.088$$

奥斯卡有 $3.8\times10^{-3}/8.8\times10^{-2}$ 即 0.043 的概率在未来的 30 天里遇到故障。

顺便说一下，一辆新购的 Itsibitsi 在 10 年里至少有一次故障的概率为 0.95。

小结

本章讨论的内容是本书的重点，我们从定义一个随机变量为从样本空间到实数直线的映射入手，然后引入了点函数 $F_X(x)$，称为累积分布函数(CDF)，它可以计算事件 $\{\zeta: \zeta \in \Omega, X(\zeta)\leqslant x\}$ 的概率。由 CDF 推导出概率密度函数(PDF)和概率质量函数(PMF)，并且讨论了许多有用的和特殊的概率分布。我们证明了如何用 δ 函数去建立离散型和连续型随机变量的统一理论。接着讨论了联合分布、泊松变换、逆泊松变换以及这些概念的应用。

我们讨论了条件概率并展示了在条件故障率中的应用，条件故障往往在开始高，在中段为常数，而在末端高，这是确定"时间－故障"的概率分布的基础。

习题

(＊星号的习题是一些高级题，可能需要更多的工作和附加的资料。)

2.1 在 n 次试验中成功 k 次(忽略次序)的事件服从二项式分布 $b(k, n; p)$。令 $n=10$, $p=0.3$。定义随机变量

$$X(k) = \begin{cases} 1, & 0\leqslant k\leqslant 2 \\ 2, & 2 < k \leqslant 5 \\ 3, & 5 < k \leqslant 8 \\ 4, & 8 < k \leqslant 10 \end{cases}$$

计算概率 $P[X=j]$, $j=1, \cdots, 4$。画出 CDF $F_X(x)=P[X\leqslant x]$(对所有的 x)。

*2.2 考虑概率空间$(\Omega, \mathcal{F}, P)$，给出一个例子，用 1～2 句话来具体化，其中所有的实验结果具有零概率。提示：根据随机变量来考虑。

2.3 一个餐厅因它与众不同的服务而著名，顾客在引起服务员注意之前要等待的时间 X(单位为分钟)服从下列 CDF：

$$F_X(x) = \begin{cases} \left(\dfrac{x}{2}\right)^2, & 0\leqslant x\leqslant 1 \\ \dfrac{x}{4}, & 1\leqslant x\leqslant 2 \\ \dfrac{1}{2}, & 2\leqslant x\leqslant 10 \\ \dfrac{x}{20}, & 10\leqslant x\leqslant 20 \\ 1, & x\geqslant 20 \end{cases}$$

(a) 画出 $F_X(x)$；(b)计算并画出 pdf $f_X(x)$，验证概率密度的面积确实为 1；(c)顾客必须等待如下时间的概率：(1)至少 10 分钟，(2)小于 5 分钟，(3)5～10 分钟之间，(4)刚好 1 分钟。

2.4 根据 $F_X(x)$ 和 $P[X=x]$(其中 $x=a, b$)，计算下列事件的概率：$\{X<a\}$、$\{X\leqslant a\}$、$\{a\leqslant X<b\}$、$\{a\leqslant X\leqslant b\}$、$\{a<X\leqslant b\}$ 和 $\{a<X<b\}$。

2.5 在下列 pdf 中，计算归一化所要求的常数 B：柯西分布($\alpha<\infty$, $\beta>0$)

$$f_X(x)=\frac{B}{1+[(x-\alpha)/\beta]^2},\qquad -\infty<x<\infty$$

麦克斯韦分布($\alpha>0$)

$$f_X(x)=\begin{cases}Bx^2\mathrm{e}^{-x^2/\alpha^2}, & x>0\\ 0, & \text{其他}\end{cases}$$

*2.6 对于下列更高级的 pdf，计算归一化所要求的常数 B:

贝塔分布($b>-1$, $c>-1$)

$$f_X(x)=\begin{cases}Bx^b(1-x)^c, & 0\leqslant x\leqslant 1\\ 0, & \text{其他}\end{cases}$$

参见参考文献[2-6]的 256 页的式(6.2-1)

χ^2分布($\sigma>0$, $n=1, 2, \cdots$)

$$f_X(x)=\begin{cases}Bx^{(n/2)-1}\exp(-x/2\sigma^2), & x>0\\ 0, & \text{其他}\end{cases}$$

2.7 噪声电阻产生的电压为 $v_n(t)$，在 $t=t_1$ 时，噪声电平 $X\triangleq v_n(t_1)$是一个高斯随机变量，pdf 为

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right]$$

计算并画出$|X|>k\sigma$($k=1, 2, \cdots$)的概率。

2.8 计算瑞利 pdf[参见式(2.4-15)]的 $F_X(k\sigma)$, $k=0, 1, 2, \cdots$。

2.9 写出伯努利分布、二项式分布和泊松 PMF 的概率密度函数(利用 δ 函数)。

2.10 随机变量的 pdf 如图 P2.10 所示。括号中的数表示面积，(a)计算 A 的值；(b)画出 CDF；(c)计算 $P[2\leqslant X<3]$；(d)计算 $P[2<X\leqslant 3]$；(e)计算 $F_X(3)$。

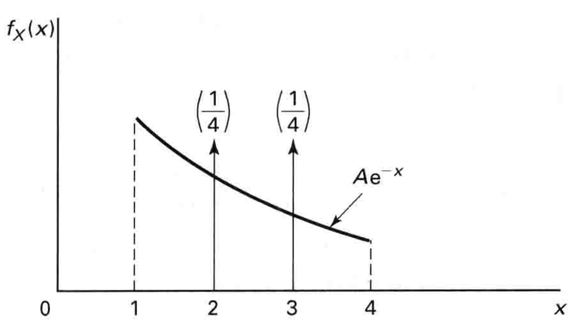


图 P2.10　混合随机变量的 pdf

2.11 随机变量 X 的 CDF 为 $F_X(x)=(1-\mathrm{e}^{-x})u(x)$，求事件$\{\zeta: X(\zeta)<1$ 或 $X(\zeta)>2\}$的概率。

2.12 随机变量 X 的 CDF 如图 P2.10 所示，括号中的数表示面积，计算 A 的值，计算 $P[2<X<4]$。

2.13 (投掷两枚硬币)随机实验是由同时投掷两枚不同的硬币组成。样本空间 $\Omega=\{\zeta_1, \zeta_2, \zeta_3\}$。定义两个随机变量

$$X(\zeta_1)=0,\ X(\zeta_2)=0,\ X(\zeta_3)=1$$
$$Y(\zeta_1)=1,\ Y(\zeta_2)=-1,\ Y(\zeta_3)=1$$

(a)计算所有可能形如 $P[\zeta: X(\zeta)=\alpha, Y(\zeta)=\beta, \alpha\in\{0, 1\}, \beta\in\{-1, 1\}$的联合概率。

(b)确定 X 和 Y 是否是独立的随机变量。

2.14 随机变量 X 的 pdf 如图 P2.14 所示，括号中的数表示对应冲激函数的面积，所以

$$f_X(x)=\frac{1}{8}\delta(x+2)+\frac{1}{16}\delta(x+1)+\frac{1}{16}\delta(x-1)+\begin{cases}Ax^2, & |x|\leqslant 2\\ 0, & \text{其他}\end{cases}$$

注意：密度 f_X 在[-2, 2]之外是零。

(a)求常数 A 的值。

(b)画出 CDF $F_X(x)$，请标出你图中有意义的点。

(c)计算 $F_X(1)$。

(d)求 $P[-1<X\leqslant 2]$。

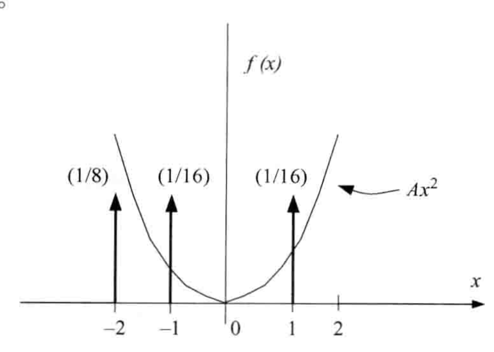


图 P2.14　习题 2.14 中混合随机变量的 pdf

2.15　考虑 PMF 为 $b(k;4,1/2)$ 的随机变量，计算 $P[X=k|X$ 为偶数$]$，$k=0,\cdots,4$。

*2.16　继续例 2.6-8，求边缘分布函数 $F_N(n)$，求并画出对应的 PMF $P_N(n)$。此外，求条件概率密度函数 $f_W(w|N=n)=f_{W|N}(w|n)$ [$f_{W|N}(w|n)$ 是给定 $N=n$ 的条件下 W 的 pdf]。

2.17　由两个工厂生产的灯泡的故障时间 X(单位为月)分别服从下列 CDF：

$$F_X(x)=(1-\mathrm{e}^{-x/5})u(x)\qquad \text{工厂 A}$$

$$F_X(x)=(1-\mathrm{e}^{-x/2})u(x)\qquad \text{工厂 B}$$

工厂 B 生产的灯泡是工厂 A 的 3 倍，灯泡是不能用肉眼分辨的，两个厂的灯泡混在一起卖。随机购买的灯泡将至少在下列时间损坏的概率：(a)2 个月；(b)5 个月；(c)7 个月。

2.18　证明 X 在给定事件 $A=\{b<X\leqslant a\}$ 条件下的条件分布为

$$F_X(x|A)=\begin{cases}0, & x\leqslant b\\ \dfrac{F_X(x)-F_X(b)}{F_X(a)-F_X(b)}, & b<x\leqslant a\\ 1, & x\geqslant a\end{cases}$$

2.19　已经证明在银行排队等待的人数 Y 服从泊松分布，即

$$P[Y=k|X=x]=\mathrm{e}^{-x}\frac{x^k}{k!},\quad k\geqslant 0,x>0$$

假定银行柜台人员的标准化服务时间 x(即为一个客户处理一笔业务的服务时间)是常数。然而，把服务时间看成随机变量更合适。为了简单起见，令 X 是均匀分布的随机变量，pdf 为

$$f_X(x)=\tfrac{1}{5}[u(x)-u(x-5)]$$

那么，$P[Y=k|X=x]$ 仍是泊松分布，但 $P[Y=k]$ 有时是其他分布。计算 $P[Y=k]$，$k=0,1,2$，对一般 k 的答案可能比较困难。

2.20　假定在总统选举中，对两个候选人(候选人 1、候选人 2)投票是等概率 $p=0.5$。假定所有投票是独立的，假定挑出 8 张选票来进行检查，令 X 表示 8 张选票中投给候选人 1 的票数，令 A 表示投向候选人 1 的票数超过 4 的事件，即 $A=\{X>4\}$。

(a) 随机变量 X 的 PMF 是多少？注意：PMF 应该是关于 $X=k=4$ 对称的。

(b) 求并画出条件分布函数 $F_X(x|A)$ $(-1\leqslant x\leqslant 10)$。

(c) 求并画出条件概率密度 $f_X(x|A)$ $(-1\leqslant x\leqslant 10)$。

(d) 求投给候选人 1 的票数在 4 ~ 5 之间的条件概率，即 $P[4\leqslant X\leqslant 5|A]$。

2.21 随机变量 X 和 Y 有联合 pdf 为

$$f_{X,Y}(x,y)=\begin{cases}\frac{3}{4}x^2(1-y),\ 0\leqslant x\leqslant 2,\ 0\leqslant y\leqslant 1\\ 0, \quad 其他\end{cases}$$

求：

(a) $P[X\leqslant 0.5]$。

(b) $F_Y(0.5)$。

(c) $P[X\leqslant 0.5|Y\leqslant 0.5]$。

(d) $P[Y\leqslant 0.5|X\leqslant 0.5]$。

2.22 考虑 X 和 Y 的联合 pdf 为

$$f_{XY}(x,y)=\frac{1}{3\pi}e^{-\frac{1}{2}[(x/2)^2+(y/3)^2]}u(x)u(y)$$

X 和 Y 统计独立吗？计算 $\{0<X\leqslant 3,\ 0<Y\leqslant 2\}$ 的概率。

2.23 考虑随机变量 X，pdf $f_X(x)$ 为

$$f_X(x)=\begin{cases}A(1+x), & -1<x\leqslant 0\\ A(1-x), & 0<x<1\\ 0, & 其他\end{cases}$$

(a)求 A 并画出 $f_X(x)$。

(b)画出 $F_X(x)$、pdf。

(c)求点 b 使得 $P[X>b]=(1/2)P[X\leqslant b]$。

2.24 证明式(2.6-75)当 $\rho=0$ 时分解为 $f_X(x)f_Y(y)$，$f_X(x)$ 和 $f_Y(y)$ 等于多少？当 $\sigma=1$ 和 $\rho=0$ 时，$P\left[-\frac{1}{2}<X\leqslant\frac{1}{2},\ -\frac{1}{2}<Y\leqslant\frac{1}{2}\right]$是多少？

*2.25 考虑一个受到噪声污染的通信信道。令 X 是发射信号的值，Y 是接收信号的值。假定在给定 $\{X=x\}$ 的条件下 Y 的条件概率密度为高斯的，即

$$f_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(y-x)^2}{2\sigma^2}\right)$$

X 在 $[-1,1]$ 上均匀分布，给定 Y 的条件下 X 的条件 pdf $f_{X|Y}(x|y)$ 等于多少？

2.26 教授到达办公室的时间是一个随机变量，在早上 8 点到 9 点之间均匀分布，定义事件

$A=\{$教授在早上 8:30 之前没有到达$\}$

$B=\{$教授将在早上 8:31 之前到达$\}$

求(a) $P[B|A]$；(b) $P[A|B]$。

2.27 令随机变量 X 的 pdf 为

$$f_X(x)=\begin{cases}0, & x<0,\\ ce^{-2x}, & x\geqslant 0,\end{cases}\qquad (c>0)$$

(a) 求 c；(b) 令 $a>0$，$x>0$，求 $P[X\geqslant x+a]$；(c) 令 $a>0$，$x>0$，求 $P[X\geqslant x+a|X\geqslant a]$。

2.28 为了庆祝通过概率论课程的考试，奈特(Wynette)邀请她的教授钱斯(Chance)博士到著名的法式餐厅共进晚餐。如果你提前 y 天预订得到座位的概率是 $1-e^{-y}$，$y\geqslant 0$。为了至少有 0.95 的概率得到座位，奈特需要提前多少天预订？

2.29 美国的防卫雷达在天空搜索不明飞行物(UFO)。令 M 表示 UFO 出现的事件，M^c 为没有 UFO。令 $f_{X|M}(x|M)=\frac{1}{\sqrt{2\pi}}\exp(-0.5[x-r]^2)$ 是当有 UFO 时雷达回波 X 的条件 pdf。令 $f_{X|M}(x|M^c)=\frac{1}{\sqrt{2\pi}}\exp(-0.5[x]^2)$ 是当无 UFO 时雷达回波 X 的条件 pdf。为了更具体，令 $r=1$，报警电平为 $x_A=0.5$。令 A 表示报警这一事件，即 $\{X>x_A\}$，计算 $P[A|M]$，$P[A^c|M]$，$P[A|M^c]$，$P[A^c|M^c]$。

2.30 在前一习题中，假定 $P[M]=10^{-3}$，计算

$$P[M|A], P[M|A^c], P[M^c|A], P[M^c|A^c]. \quad \text{对 } P[M]=10^{-6} \text{ 重复以上概率的计算}$$

注意：通过给 $P[M]$ 赋不同的数，本题试图说明在某些问题中使用概率的困难。因为 UFO 的出现是如此稀少(除了在罗斯韦尔、墨西哥)，它可以看成“一次事件”，要获得先验概率 $P[M]$ 的精确知识几乎是不可能的。因此，在 1941 年日本进攻珍珠港时，雷达清晰地指示了大规模的来袭目标云，这一信号被指挥官忽略，指挥官认为进攻的先验概率是如此之小，以至于认为最有可能的是雷达出了故障。

*2.31 (研究问题：接收机工作特性)在本题中，$P[A|M^c]$ 称为 α，即虚警概率，而 $P[M|A]$ 称为 β，即正确检测概率，很显然，$\alpha=\alpha(x_A)$，$\beta=\beta(x_A)$，写一个 MATLAB 程序，画出在固定的 r 的值的情况下 β 与 α 的图形。选 $r=0, 1, 2, 3$。在雷达界，得到的这条曲线称为不同 r 值的情况下的接收机工作特性(ROC)。

2.32 一套精致的房屋安全系统使用了红外波束来完成一个电路。如果电路断开，就说明有盗贼跨过了波束，铃声就会响。系统工作方式如下：光敏二极管按每秒 9×10^6 个光子的泊松率产生红外光子波束，计数器每一微秒对检测器收集的光子进行计数。如果计数的光子在每个计数间隔(10^{-6}秒)低于 2 个光子，就认为电路被断开，铃声响。假定泊松 PMF，计算一秒间隔内的虚警概率。

2.33 交通灯有三种状态：绿(G)、红(R)和黄(Y)，灯以随机的方式改变，在任意时间，灯只能有一种状态，考虑观察交通灯状态的随机实验。

(a) 给出这个实验的样本空间，列出它的 5 个基本元素。

(b) 令随机变量 $X(\cdot)$ 定义为：$X(G)=-1$，$X(R)=0$，$X(Y)=\pi$。假定 $P[G]=P[Y]=0.5\times P[R]$，画出 X 的 pdf，$P[X\leqslant 3]$ 等于多少？

*2.34 一种基于令牌的多用户通信系统工作如下：9 个用户站连接成一个环，一个称为令牌的电子信号按顺时针方向通过环，令牌在每个站停留并且允许用户最多有 5 分钟时间发出信息。令牌在每个站最多等待 1 分钟以便用户启动信息，如果在一分钟最后没有用户出现，令牌就传到下一个站。5 分钟窗口包括了令牌在站的等待时间，在令牌等待周期结束后开始传送信息的用户只有 4 分钟时间传送信息。

(a) 假定你是一个站的用户，你可能经历的最小和最大等待时间是多少？假定令牌从一个站传到另一个站是在瞬间完成。

(b) 假定站被占用的概率为 p，如果站被占用，占用时间在(0, 5)分钟均匀分布。用 MATLAB 写一个程序来模拟你那个站的等待时间。假定令牌刚刚离开你的站，取不同的 p 值。

2.35 令 X 和 Y 是联合高斯随机变量，pdf 为

$$f_{XY}(x, y)=\frac{1}{2\pi}\exp\left[-\frac{1}{2}(x^2+y^2)\right]$$

$P[X^2+Y^2\leqslant c^2]\geqslant 0.95$ 的最小 c 值是多少？提示：利用极坐标。

2.36 给定随机变量 X 和 Y 的联合 pdf 为(如图 P2.36 所示)

$$f_{X,Y}(x,y)=\begin{cases} A, & 0\leqslant |x|+|y|<1 \\ 0, & \text{其他} \end{cases}$$

(a) A 的值是多少？

(b) 边缘概率密度 $f_X(x)$ 是多少？

(c) X 和 Y 统计独立吗？为什么？

(d) 条件概率密度 $f_{Y|X}(y|x)$ 是多少？

2.37 假定在超市中用于扫描条形码的激光器具有恒定的条件故障率($\lambda>0$)。为了要使在 100 次使用中产生首次故障的概率小于等于 0.05，最大 λ 应该等于多少？

2.38 如果条件故障率 $\alpha(t)$ 如图 P2.38 所示，计算故障时间 X 的 pdf。

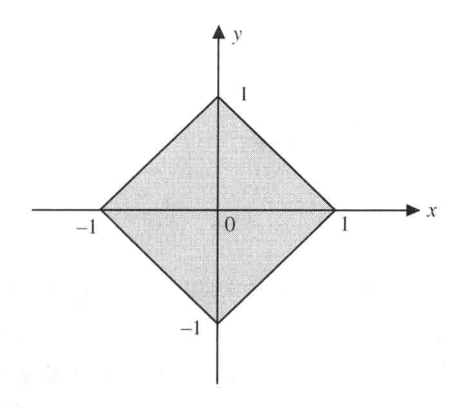


图 P2.36 习题2.36中$f_{XY}(x, y)$的支撑

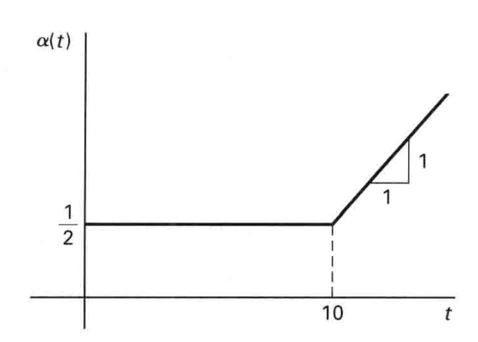


图 P2.38 习题2.38的故障率$\alpha(t)$

2.39 利用联合 CDF $F_{XY}(x, y)$的基本性质证明：

(a) $f_{XY}(x, y)\mathrm{d}x\mathrm{d}y = P[x < X \leqslant x + \mathrm{d}x, y < Y \leqslant y + \mathrm{d}y]$。

(b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\mathrm{d}x\mathrm{d}y = 1$。

(c) $f_{XY}(x, y) \geqslant 0$。

参考文献

2-1 M. Loeve, *Probability Theory.* New York: Van Nostrand, Reinhold, 1962.

2-2 W. Feller, *An Introduction to Probability Theory and Its Applications*, 2 vols. New York: John Wiley, 1950, 1966.

2-3 W. F. Davenport, *Probability and Random Processes: An Introduction for Applied Scientists and Engineers.* New York: McGraw-Hill, 1970.

2-4 B. Saleh, *Photoelectron Statistics.* New York: Springer-Verlag, 1978.

2-5 M. Born and E. Wolf, *Principles of Optics.* New York: Pergamon Press, 1965.

2-6 M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions.* New York: Dover, 1965. (On-line at various locations.)

补充文献

Cooper, G. R. and C. D. McGillem, *Probabilistic Methods of Signal and System Analysis*, 3rd edition. New York: Holt, Rinehart and Winston, 1999.

Leon-Garcia, A., *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd edition. Reading, MA: Prentice Hall, 2008.

Helstrom, C. W., *Probability and Stochastic Processes for Engineers*, 2nd edition. New York: Macmillan, 1991.

A. Papoulis, and S. U. Pillai *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, 4th Ed, 2002.

Pebbles, P. Z. Jr., *Probability, Random Variables, and Random Signal Principles*, 4th edition. New York: McGraw-Hill, 2001.

Scheaffer, R. L., *Introduction to Probability and Its Applications.* Belmont, CA: Duxbury 1990.

Yates, R. D. and D. J. Goodman, *Probability and Stochastic Processes*, 2nd edition, New York: Wiley, 2004.

Ziemer, R. E., *Elements of Engineering Probability &Statistics.* Upper Saddle River, NJ: Prentice Hall, 1997.

第 3 章　随机变量的函数

3.1　引言

在工程中有如下经典的问题：对于给定的输入我们必须计算输出结果。如果系统输入是随机的，那么输出也是随机的。说的更正式一些，如果输入在某个时刻 t 或点 x 是一个随机变量(RV)，对应时刻 t' 或点 x' 的输出也是一个随机变量。此时问题升级了，如果输入随机变量的累积分布函数(CDF)，概率质量函数(PMF)，或者概率密度函数(PDF)事先已知，那么输出随机变量的相应函数也能计算出来吗？在许多情况下是可以的，但在个别情况下计算相当困难，以至于我们只能寻求对输出随机变量性质的描述，这比累积分布函数所能提供的信息少多了。这种描述称为均值或数学期望，这部分内容将在第 4 章讨论。对于有记忆系统，计算输出的累积分布函数往往更加困难(甚至不可能)。有记忆系统是指某时刻的输出依赖于过去时刻输入的系统(有可能需要过去无限长时刻的输入值)。这就是在第 7 章和第 8 章将要讨论的随机序列和随机过程。本章讨论相对简单的情况，只涉及一个或少量随机变量。我们将采用丰富的例题来讲解。

例 3.1-1　(电阻中的功率损失)由电路理论可知，电流 I 通过电阻 R(参见图 3.1-1)产生的功率 W 由下列公式给出：

$$W \triangleq W(I) = I^2 R \tag{3.1-1}$$

根据式(3.1-1)，对于每个电流值 I 都产生对应的 $W(I)$，这种规则或联系就叫做函数。这个函数一般用 $W(\cdot)$ 或 W 来表示，有的时候也用 $W(I)$ 表示，尽管后者会使得规则和实际数值之间有些混淆。很显然，如果 I 是一个随机数，根据规则 $W = I^2R$ 将产生新的随机数 W，而且 W 的累积分布函数将会与 I 有很大差别①。实际上，这涉及一个核心问题：对于一个给定的规则 $g(\cdot)$，以及随机变量 X 和它的概率密度函数 $f_X(x)$，那么随机变量 $Y = g(X)$ 的概率密度函数 $f_Y(y)$ 是什么？

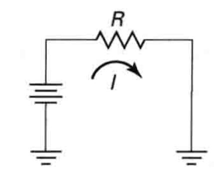


图 3.1-1　电阻中的欧姆功率损耗

计算 $f_Y(y)$，$F_Y(y)$，或者 Y 的概率质量函数(PMF) $P_Y(y_i)$，有可能很简单，也有可能很复杂。我们将用第二个例子来讲解如何计算，这个例子来自通信理论。

例 3.1-2　(检波器)由于高斯白噪声的作用，二元信号成为了模拟信号(参见图 3.1-2)。解码器在 t_0 时刻对模拟波形 $x(t)$ 进行采样，解码规则如下：

解码器输入 x	解码器输出 y
如果 $x(t_0)$：	那么 y 就被赋值为：
$\geqslant \frac{1}{2}$	1
$< \frac{1}{2}$	0

---

① 在此假设复合函数 $I^2(\zeta)R$ 满足一个随机数的性质(参见 2.2 节)。

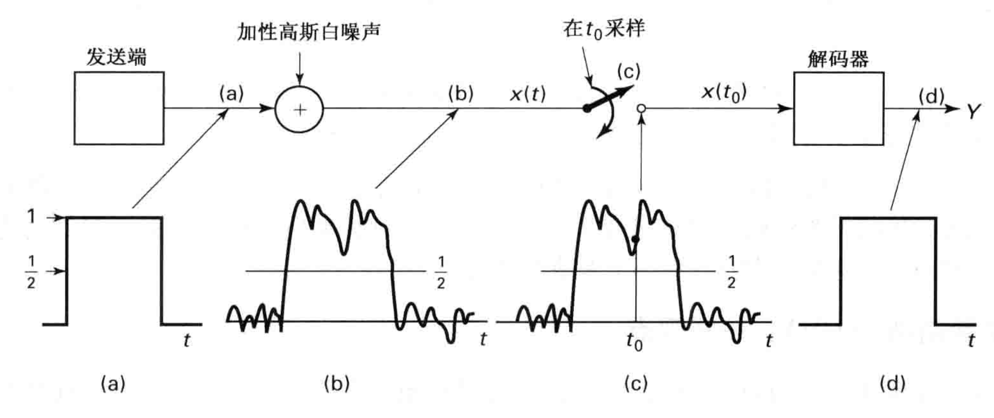


图 3.1-2　基于采样和限幅来对一个噪声污染的数字脉冲进行解码

解　由于 Y 为一个随机变量且代表了解码器的输出，我们可以写出如下的事件：

$$\{Y=0\}=\{X<0.5\} \tag{3.1-2a}$$

$$\{Y=1\}=\{X\geqslant 0.5\} \tag{3.1-2b}$$

其中 $X\triangleq x(t_0)$。如果假设 X: $N(1, 1)$，那么可得

$$P[Y=0]=P[X<0.5]=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0.5}\mathrm{e}^{-1/2(x-1)^2}\mathrm{d}x \approx 0.31 \tag{3.1-3}$$

利用 2.4 节中有关正态化过程可得到式(3.1-3)的结论，当 X: $N(0, 1)$，对于任意 $x<0$ 有 $F_X(x)=\frac{1}{2}-\mathrm{erf}(|x|)$ 成立。图 3.1-3 给出了当 $P[Y=0]$ 时，正态分布曲线 $N(1, 1)$ 下面的面积。

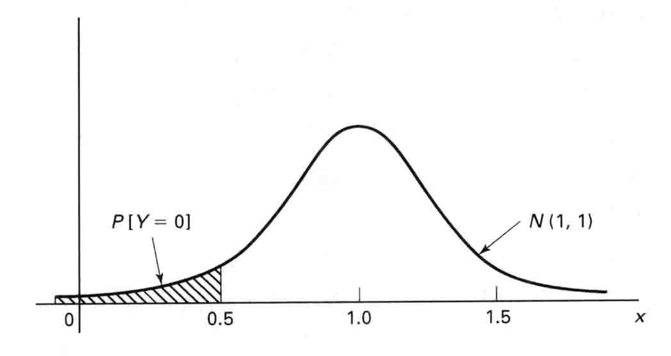


图 3.1-3　例 3.1-2 中 $P[Y=0]$ 对应的面积

同理我们可以计算出 $P[Y=1]=0.69$。因此 Y 的概率质量函数为

$$P_Y(y)=\begin{cases}0.31, & y=0\\ 0.69, & y=1\\ 0, & \text{其他}\end{cases} \tag{3.1-4}$$

利用 Dirac δ 函数，可以写出 Y 的概率密度函数表达式

$$f_Y(y)=0.31\,\delta(y)+0.69\,\delta(y-1) \tag{3.1-5}$$

Kronecker δ 函数的定义为 $y=0$ 时 $\delta(y)=1$，其余为 0。因此 Y 的概率质量函数为

$$P_Y(y)=0.31\,\delta(y)+0.69\,\delta(y-1)$$

Kronecker δ 函数常用于概率质量函数(PMF)，而 Dirac δ 函数一般用于概率密度函数(PDF)。我们用相同的符号来标记二者，但需要注意这两个函数是有区别的。

当然了，并不是所有随机变量函数(FRV)的问题都是这么简单的。为了深入了解随机变量函数问题，我们需要深刻认识随机变量函数的本质。对其本质的理解有助于我们讨论随机序列和随机过程，这部分内容将在第 7 章进行学习。

随机变量函数(FRV)：一些观点

对于一个随机变量函数，存在许多看似不同但有相同本质的观点。在此我们列举两个观点。它们之间最主要的区别就在于着眼点不同。

假设一个基本概率空间 $\mathscr{P}=(\Omega,\mathscr{F},P)$ 和定义在这个空间上的随机变量 X。X 是一个规则，可以把所有的 $\zeta\in\Omega$ 映射为一个数 $X(\zeta)$。在实数直线上，X 把事件 $\mathscr{F}$ 的 σ 域转换为实数直线上数的集合，即博雷尔的 σ 域 $\mathscr{B}$。如果 R_X 代表实数直线上的一个子空间，并且当 ζ 在 Ω 上任意取值的时候 R_X 可由 X 实现，那么可以认为 X 是定义域 Ω 和值域 R_X 上的一个普通函数。进一步地，可考虑实变量 x 的可测实函数 $g(x)$。

第一个观点 ($Y:\Omega\rightarrow R_Y$) 对于每个 $\zeta\in\Omega$，可产生一个对应的值 $g(X(\zeta))\triangleq Y(\zeta)$。对于一组随机输出 $\{\zeta\in\Omega\}$，规则 Y 可产生对应的一组数值 $\{Y(\zeta)\}$。该规则是一个随机变量，其定义域为 Ω，值域为 $R_Y\subset R^1$。对于实数 B_Y 的所有博雷尔集，集合 $\{\zeta:Y(\zeta)\in B_Y\}$ 是一个事件。实际上，事件 $\{\zeta:Y(\zeta)\leqslant y\}$ 和 $\{\zeta:g(X(\zeta))\leqslant y\}$ 是等价的。

从这个观点出发，我们着重强调 Y 是 Ω 到 R_Y 的一个映射，而不用特意考虑 X 的中间作用。

第二个观点(系统输入/输出的观点) 对于 R_X 中的每一个取值 $X(\zeta)$，我们可以生成一个新的数值 $Y=g(X)$，其取值范围为 R_Y。规则 Y 就是随机变量 X 的一个函数，其中 Y 的定义域为 R_X，值域为 R_Y。从这个观点出发，就可以把 Y 看成从一个实数集合到另外一个实数集合的映射。对该观点进行建模就可以把 X 当成一个系统的输入，该系统的传递方程为 $g(\cdot)$①。对于这样一个系统，输入 x 得到输出 $y=g(x)$，输入函数 X 得到输出函数 $Y=g(X)$［参见图 3.1-4］。

X → g(•) → Y

图 3.1-4 用输入/输出的观点来分析一个随机变量的函数

我们在此强调输入/输出的观点，这是因为在应对随机过程的时候非常有用，在这种情况下输入由随机变量的波形或序列构成。求解随机变量函数的核心问题就是：根据给定的 $g(x)$ 和 $F_X(x)$ 求出点集 C_y，这样的话，下面的事件就是等价的：

$$\begin{aligned}\{\zeta:Y(\zeta)\leqslant y\}&=\{\zeta:g[X(\zeta)]\leqslant y\}\\&=\{\zeta:X(\zeta)\in C_y\}\end{aligned}\tag{3.1-6a}$$

一般情况下，我们会采用较为简洁的写法，式(3.1-6a)在后面会被写成 $\{Y\leqslant y\}=\{X\in C_y\}$。

① g 可以是任意的可测方程。也就是说，如果 Y 的取值范围为 R_Y，那么 R_Y 中每个子集的逆像(参见 2.2 节)都是一个事件，而该逆像由 $\{Y\leqslant y\}$ 中可数的并集和交集产生。

由于 C_y 是确定的，因此有

$$P[Y \leqslant y]=P[X \in C_y] \tag{3.1-6b}$$

因为相关事件是等价的。如果 C_y 是空集，那么概率 $\{Y \leqslant y\}$ 为 0。

在处理输入输出模型时为了方便，我们常常把那些基本实验的标记省略，而直接处理随机变量 X 和 Y。对于这种方法，基本实验就成了对于 X 的观测，事件就是实数直线 R^1 上的博雷尔子集，分布函数 $F_X(\cdot)$ 取代了集合函数 $P[\cdot]$。此时 Y 就是一个映射(随机变量)，它的定义域就是 X 的值域 R_X，值域 R_Y 是 R^1 的一个子集。X 的函数特性往往被忽略了，因为我们更倾向于把 X 看成一个机制，这个机制可以产生用数值表征的随机现象。根据这个观点，X 的取值范围就无足轻重了。

关于随机变量函数的进一步讨论，可以阅读有关的参考文献①。

3.2　$Y = g(X)$ 型问题的求解

我们将讲解如何求解 $Y = g(X)$ 型的问题。最终将得到一个快速求解此类问题的公式。但是，过早地运用该公式又会掩盖问题的本质，而使得求解相当困难。

例 3.2-1　设 X 是区间(0, 1)上服从均匀分布的随机变量，即 X:$U(0, 1)$，令 $Y = 2X + 3$。此时我们需要求出式(3.1-6b)中的点集 C_y，然后再计算 $F_Y(y)$。易知

$$\{Y \leqslant y\} = \{2X + 3 \leqslant y\} = \{X \leqslant \tfrac{1}{2}(y-3)\}$$

因此 C_y 就是区间 $(-\infty, \frac{1}{2}(y-3))$，进一步有

$$F_Y(y) = F_X\left(\frac{y-3}{2}\right)$$

Y 的概率密度函数为

$$f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = \frac{\mathrm{d}}{\mathrm{d}y}\left[F_X\left(\frac{y-3}{2}\right)\right] = \frac{1}{2}f_X\left(\frac{y-3}{2}\right)$$

具体的求解过程如图 3.2-1 所示。

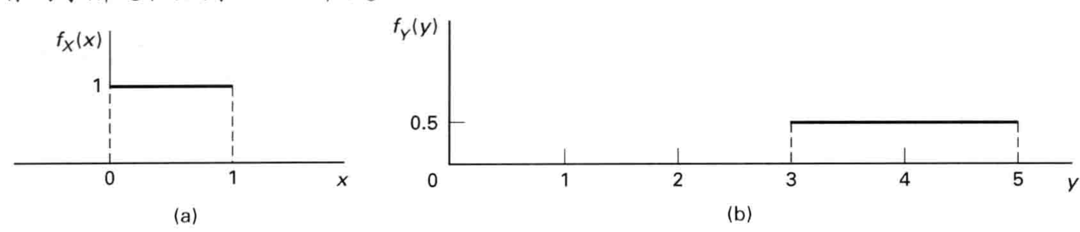


图 3.2-1　(a) X 的原始概率密度函数；(b) $Y = 2X + 3$ 的概率密度函数

一般结论　令 $Y = aX + b$，其中 X 是连续随机变量，其概率密度函数为 $f_X(x)$。当 $a > 0$ 时，输出 $\{\zeta\} \subset \Omega$ 产生的两个事件 $\{aX + b \leqslant y\}$ 与 $\{X \leqslant \frac{y-b}{a}\}$ 是等价的。那么有

$$\{Y \leqslant y\} = \{aX + b \leqslant y\} = \left\{X \leqslant \frac{y-b}{a}\right\}$$

① 参见 Davenport[3－1，p. 174]或 Papoulis、Pillai[3－5]。

根据累积分布函数的定义

$$F_Y(y)=F_X\left(\frac{y-b}{a}\right) \tag{3.2-1}$$

以及

$$f_Y(y)=\frac{1}{a}f_X\left(\frac{y-b}{a}\right) \tag{3.2-2}$$

当 $a<0$ 时，下面的事件也是等价的①

$$\{Y\leqslant y\}=\{aX+b\leqslant y\}=\left\{X\geqslant\frac{y-b}{a}\right\}$$

由于事件 $\{X<\frac{y-b}{a}\}$ 和 $\{X\geqslant\frac{y-b}{a}\}$ 是不相交的，那么它们的并集就是一个确定的事件，根据公理 3 可得

$$P\left[X<\frac{y-b}{a}\right]+P\left[X\geqslant\frac{y-b}{a}\right]=1$$

对于连续随机变量

$$P\left[X<\frac{y-b}{a}\right]=P\left[X\leqslant\frac{y-b}{a}\right]$$

以及

$$P\left[X\geqslant\frac{y-b}{a}\right]=P\left[X>\frac{y-b}{a}\right]$$

当 $a<0$ 时

$$F_Y(y)=1-F_X\left(\frac{y-b}{a}\right) \tag{3.2-3}$$

以及

$$f_Y(y)=\frac{1}{|a|}f_X\left(\frac{y-b}{a}\right),\qquad a\neq 0 \tag{3.2-4}$$

当 X 不连续时，我们需要修改 $a<0$ 的推导，$P\left[X<\frac{y-b}{a}\right]=P\left[X\leqslant\frac{y-b}{a}\right]$ 的等价性在此时可能不成立了，因为事件 $\left\{X=\frac{y-b}{a}\right\}$ 的概率可能为一个正值。修改过后的表达式成为 $P\left[X<\frac{y-b}{a}\right]=P\left[X\leqslant\frac{y-b}{a}\right]-P\left[X=\frac{y-b}{a}\right]=F_X\left(\frac{y-b}{a}\right)-P_X\left(\frac{y-b}{a}\right)$，其中利用了概率质量函数 P_X 来计算该事件的概率。$a<0$ 时的最终解答也需要做出相应的修改。

例 3.2-2 （平方律检波器）设 X 为一个连续型随机变量，其累积分布函数为 $F_X(x)$，令 $Y\triangleq X^2$。此时有

$$\{Y\leqslant y\}=\{X^2\leqslant y\}=\{-\sqrt{y}\leqslant X\leqslant\sqrt{y}\}=\{-\sqrt{y}<X\leqslant\sqrt{y}\}\cup\{X=-\sqrt{y}\} \tag{3.2-5}$$

不相交事件并集的概率就是所有概率之和。根据累积分布函数的定义，可得

---

① 在这里我们是指事件 $\{\zeta: Y(\zeta)\leqslant y\}=\left\{\zeta: X(\zeta)\geqslant\frac{y-b}{a}\right\}$。

$$F_Y(y)=F_X(\sqrt{y})-F_X(-\sqrt{y})+P[X=-\sqrt{y}] \tag{3.2-6}$$

如果 X 是连续的，$P[X=-\sqrt{y}]=0$。当 $y>0$ 时

$$f_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}[F_Y(y)]=\frac{1}{2\sqrt{y}}f_X(\sqrt{y})+\frac{1}{2\sqrt{y}}f_X(-\sqrt{y}) \tag{3.2-7}$$

当 $y<0$ 时，$f_Y(y)=0$。这个结论是如何知道的呢？利用式(3.1-6a)中的结论，当 C_y 是空集时 $P[Y\in C_y]=0$，因此有 $f_Y(y)=0$ 成立。当 $y<0$ 时，在实数直线上没有随机变量 X 满足下面的条件

$$\{-\sqrt{y}\leqslant X\leqslant\sqrt{y}\}$$

因此有 $y<0$ 时，$f_Y(y)=0$。如果 X: $N(0,1)$，根据式(3.2-7)得到

$$f_Y(y)=\frac{1}{\sqrt{2\pi y}}\mathrm{e}^{-\frac{1}{2}y}u(y) \tag{3.2-8}$$

其中 $u(y)$ 是标准单位阶跃函数。式(3.2-7)是自由度为1的 Chi 平方分布概率密度函数。

例 3.2-3　(半波检波器)半波检波器的传递方程为 $g(x)=xu(x)$，如图3.2-2所示。可得

$$F_Y(y)=P[Xu(X)\leqslant y]=\int_{\{x:\ xu(x)\leqslant y\}}f_X(x)\mathrm{d}x \tag{3.2-9}$$

(i)令 $y>0$，那么有 $\{x: xu(x)\leqslant y\}=\{x: x>0; x\leqslant y\}\cup\{x: x\leqslant 0\}=\{x: x\leqslant y\}$。

可得 $F_Y(y)=\int_{-\infty}^{y}f_X(x)\mathrm{d}x=F_X(y)$。

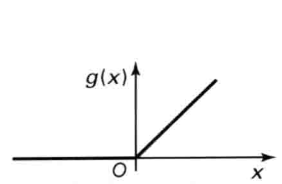
(ii)其次，令 $y=0$，那么 $P[Y=0]=P[X\leqslant 0]=F_X(0)$。

(iii)最后，令 $y<0$，有 $\{x: xu(x)\leqslant y\}=\phi$(空集)。

此时有

$$F_Y(y)=\int_{\phi}f_X(x)\mathrm{d}x=0$$

如果 X: $N(0,1)$，那么 $F_Y(y)$ 如图3.2-3所示。



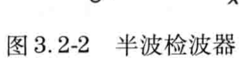


图3.2-2　半波检波器

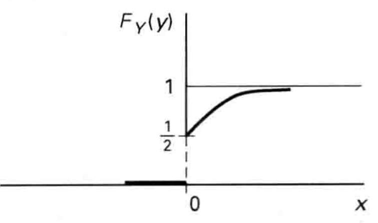


图3.2-3　X: $N(0,1)$，通过半波检波器后 Y 的累积分布函数

概率密度函数由微分得到。由于在 $y=0$ 时不连续，此时应引入 Dirac δ 函数，即

$$f_Y(y)=\begin{cases}0, & y<0\\ F_X(0)\delta(y), & y=0\\ f_X(y), & y>0\end{cases} \tag{3.2-10}$$

简洁的表达式为 $f_Y(y)=f_X(y)u(y)+F_X(0)\delta(y)$。我们注意到，$P[Y<0]=P[Y\leqslant 0]$ 在这个问题中就不成立了。此时 $P[Y=0]$ 就是一个非0的概率值。

例 3.2-4　设 X 为服从伯努利分布的随机变量，其中 $P[X=0]=p$，$P[X=1]=q$。那么有

$$f_X(x)=p\delta(x)+q\delta(x-1) \quad 和 \quad F_X(x)=pu(x)+qu(x-1)$$

其中 $u(x)$ 是连续变量 x 的单位阶跃函数。

令 $Y \triangleq X-1$，可以得到(参见图 3.2-4)

$$\begin{aligned} F_Y(y) &= P[X-1 \leqslant y] \\ &= P[X \leqslant y+1] \\ &= F_X(y+1) \\ &= pu(y+1)+qu(y) \end{aligned}$$

概率密度函数为

$$f_Y(y)=\frac{\mathrm{d}}{\mathrm{d}y}[F_Y(y)]=p\delta(y+1)+q\delta(y) \tag{3.2-11}$$

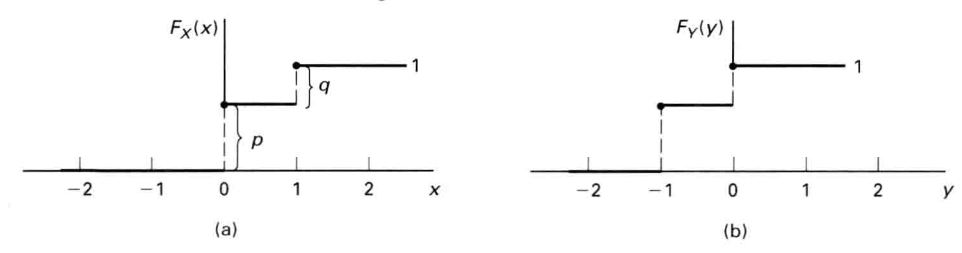


图 3.2-4 (a) X 的累积分布函数；(b) $Y=X-1$ 的累积分布函数

例 3.2-5 (累积分布函数的转换)设 X 具有连续的累积分布函数 $F_X(x)$，且 $F_X(x)$ 是 x 的严格单调增函数①。设 Y 是一个随机变量，通过 X 的累积分布函数来生成

$$Y=F_X(X) \tag{3.2-12}$$

为了计算 $F_Y(y)$，我们按照通常的步骤可得

$$\begin{aligned} \{Y \leqslant y\} &= \{F_X(X) \leqslant y\} \\ &= \{X \leqslant F_X^{-1}(y)\} \end{aligned}$$

因此

$$\begin{aligned} F_Y(y) &= P[F_X(X) \leqslant y] \\ &= P[X \leqslant F_X^{-1}(y)] \\ &= \int_{\{x:\ F_X(x) \leqslant y\}} f_X(x)\mathrm{d}x \end{aligned}$$

1. 令 $y<0$。因为对所有的 $x \in [-\infty, \infty]$ 都有 $0 \leqslant F_X(x) \leqslant 1$，集合 $\{x: F_X(x) \leqslant y\}=\phi$，所以有 $F_Y(y)=0$。
2. 令 $y>1$。此时 $\{x: F_X(x) \leqslant y\}=[-\infty, \infty]$，所以 $F_Y(y)=1$。
3. 令 $0 \leqslant y \leqslant 1$，此时 $\{x: F_X(x) \leqslant y\}=\{x: x \leqslant F_X^{-1}(y)\}$。
   所以可以得到

$$F_Y(y)=\int_{-\infty}^{F_X^{-1}(y)} f_X(x)\mathrm{d}x=F_X(F_X^{-1}(y))=y$$

最终解为

① 换句话说，当 $x_2>x_1$ 时 $F_X(x_2)>F_X(x_1)$，不会出现二者相等的情况。

$$F_Y(y)=\begin{cases}0, & y<0\\ y, & 0\leqslant y\leqslant 1\\ 1, & y>1\end{cases} \tag{3.2-13}$$

通过式(3.2-13)可知：不管X服从什么分布，只要它的累积分布函数连续且严格单调，那么$Y\triangleq F_X(X)$服从均匀分布。反过来，给定一个服从均匀分布的Y，通过变换$X\triangleq F_X^{-1}(Y)$可以产生一个随机变量X，该随机变量的累积分布函数连续且严格单调(如图3.2-5所示)。有的时候可以利用均匀分布的随机变量，通过这种方法来模拟产生特定分布的随机变量。

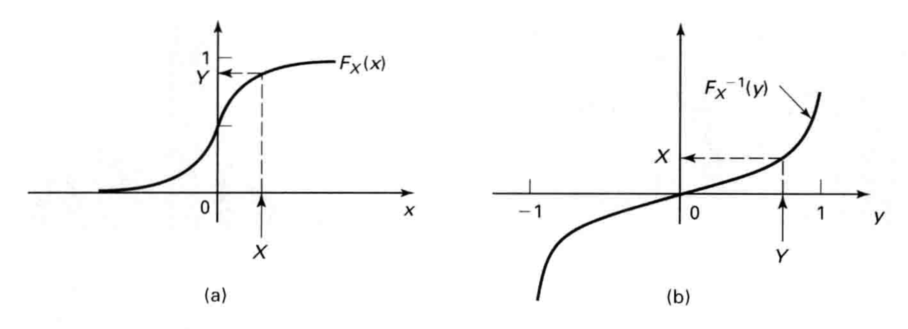


图3.2-5　通过均匀分布随机变量产生一个累积分布函数为$F_X(x)$的随机变量。(a)通过随机变量X产生随机变量Y，X的累积分布函数为$F_X(x)$，Y服从均匀分布；(b)通过随机变量Y产生随机变量X，Y服从均匀分布，X的累积分布函数为$F_X(x)$

例3.2-6　(均匀分布转换为标准正态分布)从上个例子可知，我们可以把一个均匀分布的随机变量X: $U[0, 1]$转换为任意分布的连续随机变量，转换后的累积分布函数严格递增。如果希望得到一个标准正态分布Y: $N(0, 1)$，其累积分布函数为

$$F_Y(y)=\begin{cases}\frac{1}{2}+\operatorname{erf}(y), y\geqslant 0\\ \frac{1}{2}-\operatorname{erf}(-y), y<0\end{cases}$$

求解传递函数$g(x)$可得

$$g(x)=\begin{cases}-\operatorname{erf}^{-1}(\frac{1}{2}-x), 0\leqslant x<\frac{1}{2}\\ \operatorname{erf}^{-1}(x-\frac{1}{2}), \frac{1}{2}\leqslant x\leqslant 1\end{cases}$$

传递函数如图3.2-6所示。

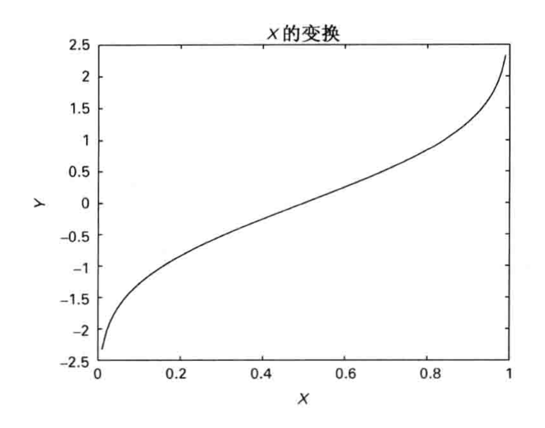


图3.2-6　$y=g(x)=F_{\text{SN}}^{-1}(x)$的变换图，可把$U[0, 1]$转换为$N(0, 1)$

在本书的网站上，有 transformCDF.m 的 MATLAB 程序可供下载，运行该程序可产生相对频率直方图。我们运行了 1000 次仿真实验得到接下来的结果。图 3.2-7 给出了 1000 个 [0, 1]之间均匀分布随机变量的直方图。图 3.2-8① 给出了转换后随机变量的直方图。

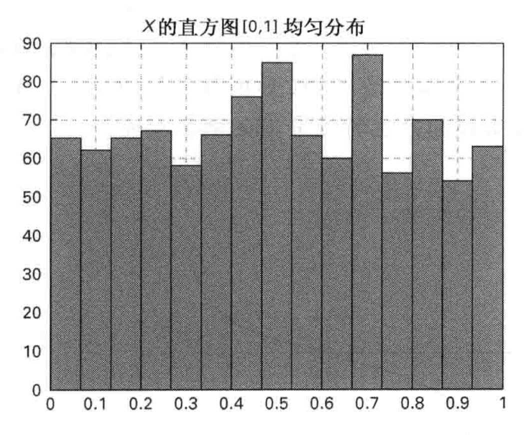


图 3.2-7 1000 个独立同分布随机变量的直方图，服从 $U[0,1]$

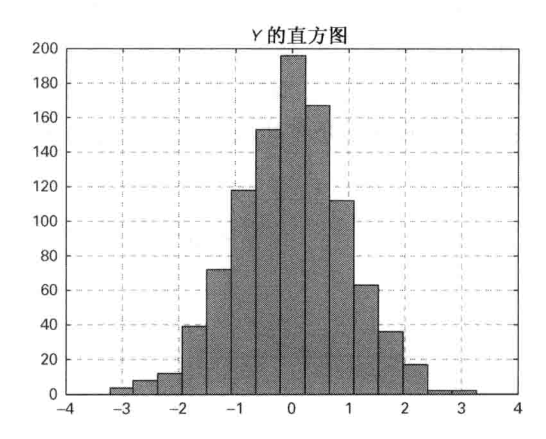


图 3.2-8 1000 个转换后的独立同分布随机变量的直方图

例 3.2-7 (量化)在一个模数转换系统中，模拟波形将被采样，量化和编码(如图 3.2-9 所示)。量化器就是一个函数，它把集合 $Q \triangleq \{y_{-N}, \cdots, y_0, \cdots, y_N\}$ 中的数值赋给每个样本 x，集合中 $2N+1$ 个值都是预先给定的[3-2]。于是，一个无限不可数的集合(模拟输入 x)就被缩减为一个有限的集合(数字输出 y_i)。注意到这个量化器其实也是一个限幅器，对于所有大于 y_N 或者小于 y_{-N} 的 x，输出 y 都被限制为 y_N 或 y_{-N}。

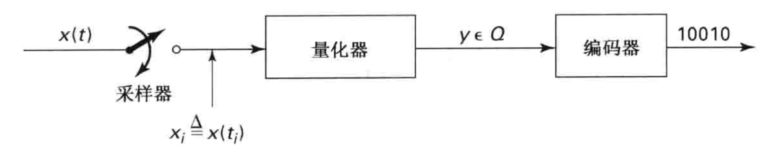


图 3.2-9 模数转换器

一个常用的量化器就是均匀量化器，该量化器是一个具有均匀步长 a 的阶梯函数，表达式为

$$g(x) = ia \qquad (i-1)a < x \leqslant ia, \qquad i \text{ 为整数} \tag{3.2-14}$$

此时，量化器将连续样本上方最接近 x 值的 ia 赋给 x②，如图 3.2-10 的阶梯函数所示。

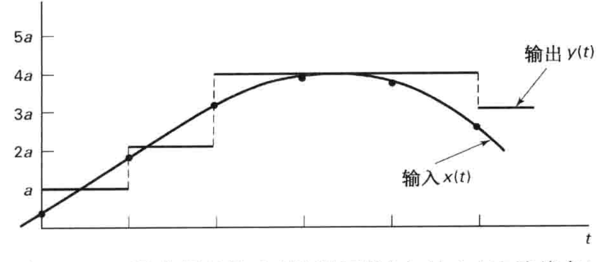


图 3.2-10 量化器的输出(阶梯函数)与输入(连续线条)

---

① 原文在这里是图 3.2-8(b)，疑有误——译者注。

② 原文为 $3x$，疑有误——译者注。

如果 X 代表样本输入值，Y 代表量化输出，且 $a=1$。X 是一个随机变量，那么输出的概率质量函数为

$$\begin{aligned}P_Y(i) &= P[Y=i]\\ &= P[i-1<X\leqslant i]\\ &= F_X(i)-F_X(i-1)\end{aligned}$$

累积分布函数此时也变成了阶梯函数，如图 3.2-11 所示①。

$$\begin{aligned}F_Y(y) &= \sum_i P_Y(i)u(y-i)\\ &= \sum_i [F_X(i)-F_X(i-1)]u(y-i)\end{aligned} \tag{3.2-15}$$

如果 $y=n$（n 为整数），$F_Y(y)=F_X(y)$，否则 $F_Y(y)<F_X(y)$。

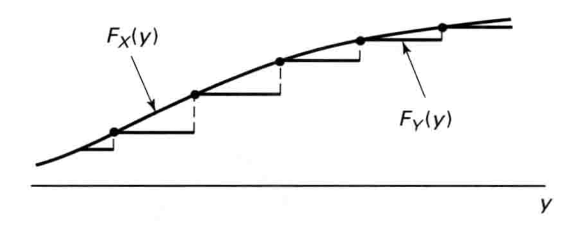


图 3.2-11　$F_X(y)$ 与 $F_Y(y)$

例 3.2-8　（正弦波）计算 $Y=\sin X$ 的概率密度函数是一个非常经典的问题，其中 $X:U(-\pi,+\pi)$ 表示区间 $(-\pi,+\pi)$ 上均匀分布的随机变量。从图 3.2-12 可知，对于所有的 $0\leqslant y\leqslant 1$，事件 $\{Y\leqslant y\}$ 满足如下等式：

$$\begin{aligned}\{Y\leqslant y\} &= \{\sin X\leqslant y\}\\ &= \{-\pi<X\leqslant \arcsin y\}\cup\{\pi-\arcsin y<X\leqslant\pi\}\end{aligned}$$

因为后面两个事件是不相交的，我们可以得到

$$F_Y(y)=F_X(\pi)-F_X(\pi-\arcsin y)+F_X(\arcsin y)-F_X(-\pi) \tag{3.2-16}$$

因此

$$\begin{aligned}f_Y(y) &= \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y}\\ &= f_X(\pi-\arcsin y)\frac{1}{\sqrt{1-y^2}}+f_X(\arcsin y)\frac{1}{\sqrt{1-y^2}}\end{aligned} \tag{3.2-17}$$

$$=\frac{1}{2\pi}\frac{1}{\sqrt{1-y^2}}+\frac{1}{2\pi}\frac{1}{\sqrt{1-y^2}} \tag{3.2-18}$$

$$=\frac{1}{\pi}\frac{1}{\sqrt{1-y^2}},\quad 0\leqslant y<1 \tag{3.2-19}$$

如果把上述计算过程在 $-1<y\leqslant 0$ 时再进行一次，图 3.2-12 的结果就变为图 3.2-13。此时以随机变量 X 来描述事件 $\{Y\leqslant y\}=\{\sin X\leqslant y\}$ 得到

① 注：原文为图 3.2-12，疑有误——译者注。

$$\{-\pi-\arcsin y<X\leqslant \arcsin y\} \tag{3.2-20}$$

这里用到了 $y\leqslant 0$ 时的 sin 函数的逆函数。此时我们可以写出累积分布函数的具体表达式

$$F_Y(y)=F_X(\arcsin y)-F_X(-\pi-\arcsin y)$$

通过微分可得

$$\begin{aligned}f_Y(y)&=\frac{\mathrm{d}F_Y(y)}{\mathrm{d}y}\\&=f_X(\arcsin y)\frac{1}{\sqrt{1-y^2}}-f_X(-\pi-\arcsin y)\frac{1}{\sqrt{1-y^2}}\\&=\frac{1}{2\pi}\frac{1}{\sqrt{1-y^2}}+\frac{1}{2\pi}\frac{1}{\sqrt{1-y^2}}\\&=\frac{1}{\pi}\frac{1}{\sqrt{1-y^2}},\qquad -1<y\leqslant 0\end{aligned}$$

与 $0\leqslant y<1$ 时的结果表达式是相同的。

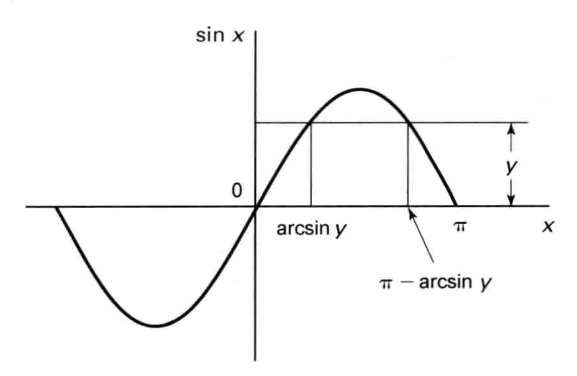


图 3.2-12 $y=\sin x$ 的根，$0\leqslant y<1$

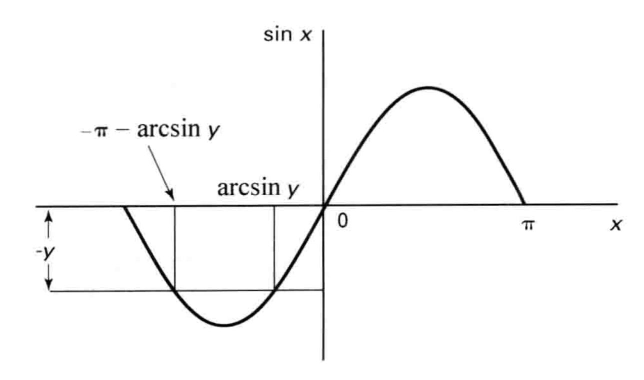


图 3.2-13 $y=\sin x$ 的根，$-1<y<0$

最后考虑 $|y|>1$ 的情况，由于对于所有的 x 都有 $|\sin(x)|\leqslant 1$，因此此时概率密度函数 f_Y 一定为 0。归纳上面得到的三个结果，完整的结果如下(参见图 3.2-14)：

$$f_Y(y)=\begin{cases}\dfrac{1}{\pi}\dfrac{1}{\sqrt{1-y^2}}, & |y|<1\\ 0, & \text{其他}\end{cases} \tag{3.2-21}$$

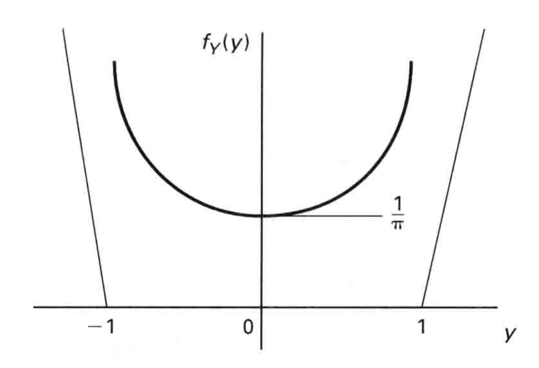


图 3.2-14　$Y = \sin X$ 的概率密度函数

接下来，我们将推导一种求解 $Y = g(X)$ 概率密度函数的简单公式。利用这种公式，可以从概率密度函数直接得到变换后的概率密度函数，而无须先求累积分布函数。这种方法称为直接法。但是对于有些问题，用间接法会减小犯错的可能，这将在后面的章节讲解。

求解 $Y = g(X)$ 概率密度函数的一般公式

对于一个连续随机变量 X，其概率密度函数为 $f_X(x)$，实变量 x 存在可微函数 $g(x)$。求 $Y \triangleq g(X)$ 的概率密度函数。

解　事件 $\{y < Y \leqslant y + \mathrm{d}y\}$ 可以被写成在博雷尔域上一系列不相交基本事件 $\{E_i\}$ 的并集。如果方程 $y = g(x)$ 有 n 个实根①$x_1, \cdots, x_n$，那么不相交事件 $E_i = \{x_i - |\mathrm{d}x_i| < X \leqslant x_i\}$ ($g'(x_i)$ 为负或者 $E_i = \{x_i < X \leqslant x_i + |\mathrm{d}x_i|\}$ ($g'(x_i)$ 为正②，如图 3.2-15 所示。对于任意一种情况，通过概率密度函数的定义可知 $P[E_i] = f_X(x_i)|\mathrm{d}x_i|$，因此有

$$\begin{aligned} P[y < Y \leqslant y + \mathrm{d}y] &= f_Y(y)|\mathrm{d}y| \\ &= \sum_{i=1}^{n} f_X(x_i)|\mathrm{d}x_i| \end{aligned} \tag{3.2-22}$$

等式两边同时除以 $|\mathrm{d}y|$ 可得

$$f_Y(y) = \sum_{i=1}^{n} f_X(x_i)\left|\frac{\mathrm{d}x_i}{\mathrm{d}y}\right| = \sum_{i=1}^{n} f_X(x_i)\left|\frac{\mathrm{d}y}{\mathrm{d}x_i}\right|^{-1}$$

在方程 $y = g(x)$ 根的位置，有 $\mathrm{d}y/\mathrm{d}x_i = g'(x_i)$，可得下面重要的公式：

$$f_Y(y) = \sum_{i=1}^{n} f_X(x_i)/|g'(x_i)| \qquad x_i = x_i(y), \qquad g'(x_i) \neq 0 \tag{3.2-23}$$

当传递函数 $g(x)$ 有多个根时，式(3.2-23)在求解此类问题时非常有用，在此假设对所有根的求导 $g'(x_i) \neq 0$。我们可以看看 $g'(x_i) = 0$ 会是什么情况，导数为 0 意味着在这个区域函数 g 是平坦的。对于该平坦区域的任何 x 值，对应的 y 都是一个常数，在这个区域 y 取值的概率质量就等于事件 X 发生在该平坦区域的概率。从 f_Y 的角度，此时概率质量就变为了一个冲激，且冲激的大小等于概率质量。

---

① 根是指满足 $y - g(x_i) = 0$，$i = 1, \cdots, n$ 的点 x_i 的集合。

② g 上方的一撇是指对 x 求导。

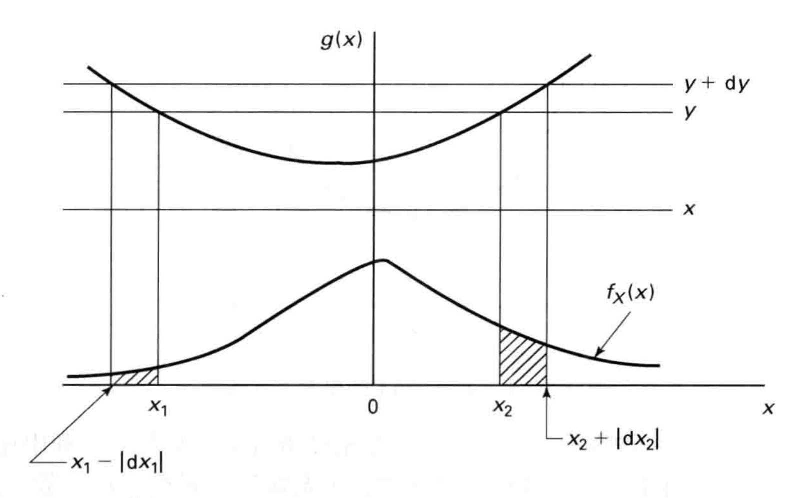


图3.2-15 事件$\{y < Y \leqslant y + \mathrm{d}y\}$是概率空间$X$上两个不相交事件的并集

如果对于一个给定的y，方程$y - g(x) = 0$没有实根，那么此时$f_Y = 0$[①]。图3.2-15给出了$n = 2$时的情况。

例3.2-9 （X的三角函数）为演示如何使用式(3.2-23)，我们在此用该公式来求解例3.2-8。已知X的概率密度函数为$f_X(x) = 1/2\pi$，$-\pi < x \leqslant \pi$，求$Y = \sin X$的概率密度函数。在这里传递函数g为$g(x) = \sin(x)$。当$y > 0$时，方程$y - g(x) = y - \sin x = 0$的两个根分别为$x_1 = \arcsin y$和$x_2 = \pi - \arcsin y$。

$$\frac{\mathrm{d}g}{\mathrm{d}x} = \cos x$$

需要计算上式在根x_1和x_2处的值。对于$x_1 = \arcsin y$，可得$\mathrm{d}g/\mathrm{d}x|_{x=x_1} = \cos(\arcsin y)$。同样对于根$x_2 = \pi - \arcsin y$，可以得到

$$\left.\frac{\mathrm{d}g}{\mathrm{d}x}\right|_{x=x_2} = \cos(\pi - \arcsin y) = \cos\pi \times \cos(\arcsin y) + \sin\pi \times \sin(\arcsin y)$$

$$= -\cos(\arcsin y)$$

利用图3.2-16的帮助，可以求出$\cos(\arcsin y)$的值。根据$\theta = \arcsin y$和$\cos\theta = \sqrt{1 - y^2} = \cos(\arcsin y)$，可以得到

$$\left|\left.\frac{\mathrm{d}g}{\mathrm{d}x}\right|_{x_1}\right| = \left|\left.\frac{\mathrm{d}g}{\mathrm{d}x}\right|_{x_2}\right| = \sqrt{1 - y^2}$$

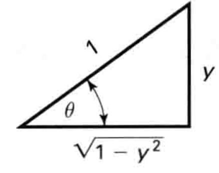


图3.2-16 计算$\cos(\arcsin y)$

最后，$f_X(\arcsin y) = f_X(\pi - \arcsin y) = 1/2\pi$。利用式(3.2-23)中的结论可得

$$f_Y(y) = \frac{1}{\pi}\frac{1}{\sqrt{1 - y^2}}, \qquad 0 \leqslant y < 1$$

结果与式(3.2-19)一致。当$y < 0$时，重复上面的计算过程，最终结果与式(3.2-21)一致。

例3.2-10 （非线性元器件）一些无记忆的非线性元器件可以用传递函数$g(x) = x^n$来进

① 对于实值的随机变量X，是不可能出现虚数的。

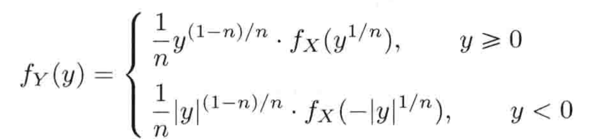
行建模。令 $Y = X^n$。Y 的概率密度函数取决于 n 是偶数还是奇数。我们在此求解 n 为奇数的情况，n 为偶数的情况留做习题。当 n 为偶数且 $y > 0$ 时，方程 $y - x^n = 0$ 只有唯一的实根 $x_r = y^{1/n}$，可得

$$\frac{\mathrm{d}g}{\mathrm{d}x} = nx^{n-1} = ny^{(n-1)/n}$$

当 $y < 0$ 时，唯一的实根为 $x_r = -|y|^{1/n}$，如图 3.2-17 所示。类似的有

$$\frac{\mathrm{d}g}{\mathrm{d}x} = n|y|^{(n-1)/n}$$

综上所述



$$f_Y(y) = \begin{cases} \frac{1}{n} y^{(1-n)/n} \cdot f_X(y^{1/n}), & y \geqslant 0 \\ \frac{1}{n} |y|^{(1-n)/n} \cdot f_X(-|y|^{1/n}), & y < 0 \end{cases}$$

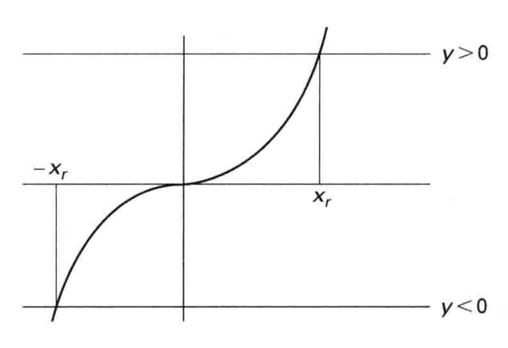


图 3.2-17　$g(x) = x^n - y = 0$ 的根，n 为奇数

如果在这道题目中 $g(x)$ 为一个常数，即 $g(x) = c$，在某些宽度非 0 的区间就不能套用式(3.2.23)来计算 $f_Y(y)$，因为在此区间上 $g'(x) = 0$。此时就需要额外考虑这个平坦区域引起概率质量的值。

例 3.2-11　(具有截止值的线性放大器)考虑图 3.2-18 所示的非线性元器件，其传递函数 $g(x)$ 如下：

$$g(x) = 0, \quad |x| \geqslant 1 \tag{3.2-24}$$

$$g(x) = x, \quad -1 < x < 1 \tag{3.2-25}$$

可以看出，当 $|x| \geqslant 1$ 时 $g'(x) = 0$，$-1 < x < 1$ 时 $g'(x) = 1$。当 $y \geqslant 1$ 和 $y \leqslant -1$ 时，方程 $y - g(x) = 0$ 没有实根，因此在这个区间上 $f_Y(y) = 0$。当 $-1 < y < 1$ 时，方程 $y - g(x) = y - x = 0$ 唯一的实根就是 $x = y$，在这个区间上运用式(3.2-23)，且有 $|g'(x)| = 1$，所以 $f_Y(y) = f_X(y)$。需要注意 $P[Y = 0] = P[X \geqslant 1] + P[X \leqslant -1]$。如果 X: $N(0, 1)$，那么 $P[X \geqslant 1] = 1/2 - \mathrm{erf}(1) = P[X \leqslant -1]$，$P[Y = 0] = 1 - 2\mathrm{erf}(1) = 0.317$。我们习惯把 $P[Y = 0] = 0.317$ 的结果代入 Y 的概率密度函数表达式，这样就需要引入 δ 函数

$$P[Y = 0] = 0.317 = \lim_{\epsilon \to 0} \int_{0-\epsilon}^{0+\epsilon} 0.317\delta(y)\mathrm{d}y$$

代入 $0.317\delta(y)$ 的表达式到 $f_Y(y)$ 中，得到最终的结果如下：

$$f_Y(y) = \begin{cases} 0, & |y| \geqslant 1 \\ (2\pi)^{-1/2} \exp\left(-\frac{1}{2}y^2\right) + 0.317\delta(y), & -1 < y < 1 \end{cases}$$

例 3.2-12　(无限多个根)在此我们考虑将图 3.2-18 的传递函数进行周期拓展，拓展后的传递函数 $g(x)$ 如图 3.2-19 所示。

在这种情况下传递函数的表达式为

$$g(x) = \sum_{n=-\infty}^{\infty} (x - 2n) \operatorname{rect}\left(\frac{x - 2n}{2}\right)$$

在这里 rect 是一个对称的单位脉冲函数，定义如下：

$$\mathrm{rect}(x)=\begin{cases}1, & -0.5<x<+0.5\\ 0, & \text{其他}\end{cases}$$

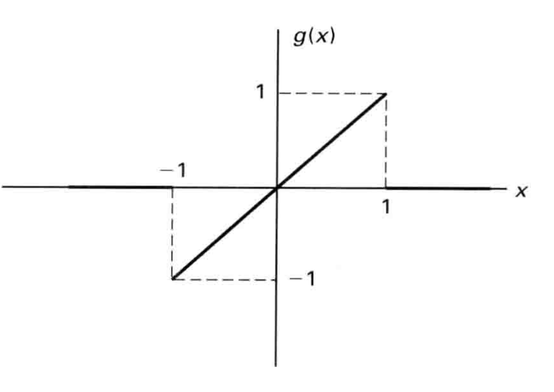


图 3.2-18 具有截止值的线性放大器

同上一个例题类似，$|y|\geqslant 1$ 时 $f_Y(y)=0$ 这是因为在这个区间方程 $y-g(x)=0$ 没有实根。另一方面，当 $-1<y<1$ 时，方程 $y-g(x)=0$ 有无限多个根，具体为 $x_n=y+2n$，$n=\cdots,-2,-1,0,1,2,\cdots$。对于每个根都有 $|g'(x_n)|=1$。因此，根据式(3.2-23)可得 $f_Y(y)=\sum_{n=-\infty}^{\infty}f_X(y+2n)\mathrm{rect}(\frac{y}{2})$。如果 X: $N(0,1)$

$$f_Y(y)=(2\pi)^{-1/2}\sum_{n=-\infty}^{\infty}\exp\left(-\frac{1}{2}(y+2n)^2\right)\times\mathrm{rect}\left(\frac{y}{2}\right)$$

如果这个结果是正确的，似乎很难理解为什么由无限多个正值之和组成的方程，其面积为 1。为了验证 $f_Y(y)$ 积分值确实为 1，我们进行如下的推导

$$\int_{-\infty}^{\infty}f_Y(y)\mathrm{d}y=\frac{1}{\sqrt{2\pi}}\sum_{n=-\infty}^{\infty}\int_{-1}^{1}\exp\left(-\frac{1}{2}(y+2n)^2\right)\mathrm{d}y \tag{3.2-26}$$

$$=\frac{1}{\sqrt{2\pi}}\sum_{n=-\infty}^{\infty}\int_{-1+2n}^{1+2n}\exp\left(-\frac{1}{2}y^2\right)\mathrm{d}y \tag{3.2-27}$$

$$=\sum_{n=-\infty}^{\infty}[\mathrm{erf}(1+2n)-\mathrm{erf}(-1+2n)] \tag{3.2-28}$$

把上面最后一项展开，读者很快就会发现除了第一项（$n=-\infty$）和最后一项（$n=\infty$）外，其余项都可以互相抵消。因此可得

$$\int_{-\infty}^{\infty}f_Y(y)\mathrm{d}y=\mathrm{erf}(\infty)-\mathrm{erf}(-\infty)=2\times\mathrm{erf}(\infty)=1$$

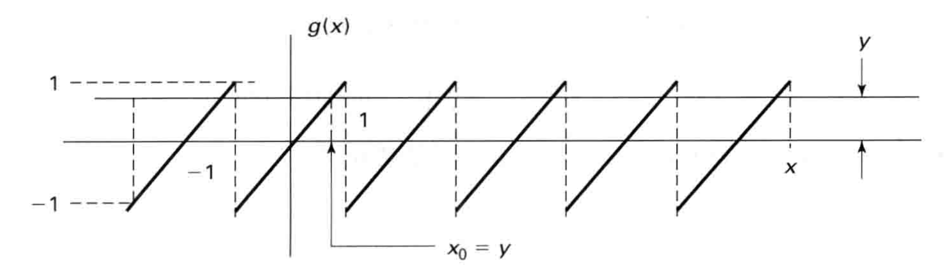


图 3.2-19 周期的传递函数

3.3 $Z=g(X,Y)$ 型问题的求解

在许多科学和工程问题中，随机变量 Z 往往是两个(或多个)随机变量 X, Y 的函数。例如：

1. 放大器的输入信号 Z 是由信号 X 和独立随机噪声 Y 叠加而成。因此 $Z = X + Y$，如果 X 也是一个随机变量，那么 Z 的概率密度函数是什么(参见图 3.3-1)?

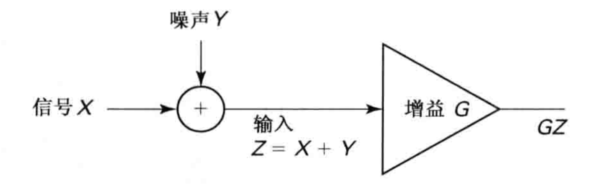


图 3.3-1　信号与独立加性噪声相加的问题

2. 只要至少有一个引擎在工作，双引擎飞机就能飞行。用 X 和 Y 来分别表示飞机右引擎和左引擎停止工作的时间，那么飞机失事的时间为 $Z \triangleq \max(X, Y)$。试求 Z 的概率密度函数?
3. 许多信号处理系统会将两个信号相乘(调制器，解调器，相关器，等等)。如果 X 是一个输入信号，Y 是另外一个输入信号，试求输出信号 $Z \triangleq XY$ 的概率密度函数。
4. 许多物理问题中经常遇到著名的“随机游走”问题，一个粒子在 x 和 y 方向的位移是独立的随机变量 X 和 Y。试求总的位移 $Z \triangleq [X^2 + Y^2]^{1/2}$ 的概率密度函数(参见图 3.3-2)。

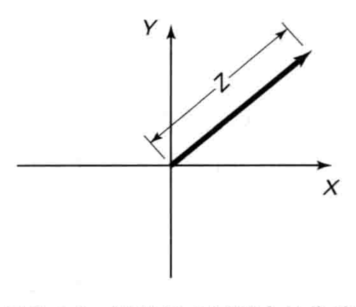


图 3.3-2　随机游走问题中的位移

$Z = g(X, Y)$ 类型的问题与 3.2 节中讨论的问题并没有本质的区别。对于 $Y = g(X)$，最基本的问题就是要找出点集 C_y，使得事件 $\{\zeta: Y(\zeta) \leqslant y\}$ 与 $\{\zeta: X(\zeta) \in C_y\}$ 等价。在这里其实也是要解答同样的问题：在(x, y)平面上找出点集 C_z，使得事件 $\{\zeta: Z(\zeta) \leqslant z\}$ 和 $\{\zeta: X(\zeta), Y(\zeta) \in C_z\}$ 是等价的，我们常常简记为

$$\{Z \leqslant z\} = \{(X,Y) \in C_z\} \tag{3.3-1}$$

以及

$$F_Z(z) = \iint_{(x,y)\in C_z} f_{XY}(x,y)\mathrm{d}x\,\mathrm{d}y \tag{3.3-2}$$

点集 C_z 由函数 $g(x, y) \leqslant z$ 决定。很显然，处理 $Z = g(X, Y)$ 类型的问题将会用联合密度，联合分布和双重积分(或求和)来取代单一的变量。此时，计算 $f_Z(z)$ 应该比计算 $Y = g(X)$ 中的 $f_Y(y)$ 复杂。我们后面将会学到两个非常省时省力的方法。(1)我们有一个“转动把柄”的公式来解决许多 $Y = g(X, Y)$ 类型的问题，这实际上是式(3.2-23)的一个扩展，此时需要用到一个辅助变量(参见 3.4 节)；(2)利用特征函数可以解决 $Z = X + Y$ 类型的问题。但是，应用这些简易方法容易混淆对本质概念的理解。

现在，我们将用前面给出的那些原理来求解下面的问题。

例 3.3-1　(随机变量的乘积)根据式(3.3-2)，计算 $Z = XY$ 的累积分布函数需要求点集 C_z，即求出满足 $g(x, y) \triangleq xy \leqslant z$ 的区域。当 $z > 0$ 时，该区域如图 3.3-3 所示。

根据图形，可计算如下：

$$F_Z(z) = \int_0^{\infty}\left(\int_{-\infty}^{z/y} f_{XY}(x,y)\mathrm{d}x\right)\mathrm{d}y + \int_{-\infty}^{0}\left(\int_{z/y}^{\infty} f_{XY}(x,y)\mathrm{d}x\right)\mathrm{d}y,\quad z \geqslant 0 \tag{3.3-3}$$

为了计算概率密度f_Z，需要对等式关于z求导。我们可以直接对式(3.3-3)进行求导运算，不过，为了对该过程认识得更清晰一些，我们首先定义不定积分$G_{XY}(x, y)$

$$G_{XY}(x,y) \triangleq \int f_{XY}(x,y)\mathrm{d}x \tag{3.3-4}$$

那么有

$$F_Z(z) = \int_0^{\infty} [G_{XY}(z/y,y) - G_{XY}(-\infty,y)]\mathrm{d}y + \int_{-\infty}^{0} [G_{XY}(\infty,y) - G_{XY}(z/y,y)]\mathrm{d}y$$

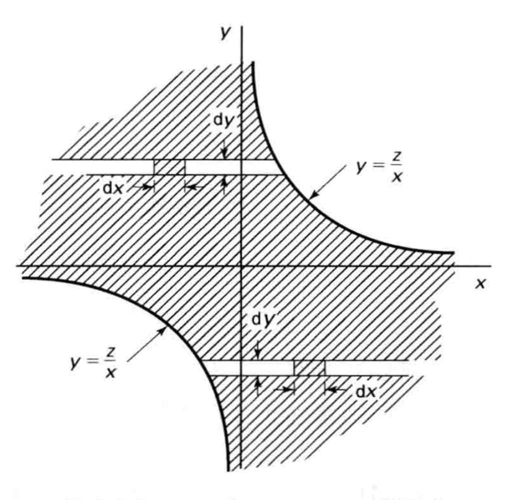


图3.3-3　$z>0$时，$xy \leqslant z$的区域

对等式两边关于z求导可得

$$f_Z(z) = \frac{\mathrm{d}F_Z(z)}{\mathrm{d}z} = \int_{-\infty}^{\infty} \frac{1}{|y|} f_{XY}(z/y,y)\mathrm{d}y \tag{3.3-5}$$

利用附录A中的式(A2-1)，对式(3.3-3)的z直接求导也可以得到相同的结果。

现在的问题就是求当$z<0$时的结果。当$z<0$时式(3.3-3)是成立的[①]，因此对于所有的z也是成立的。$z<0$的情况如图3.3-4所示。通过该图可知，当$z<0$时在阴影区域$\{xy \leqslant z\}$的积分结果$F_Z(z)$与式(3.3-3)完全相同。对该式关于z求导，再把y移到积分表达式里面，我们再次得到式(3.3-5)[②]。综上可得两个随机变量乘积的概率密度函数的一般表达式

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_{XY}(z/y,y)\mathrm{d}y, \quad -\infty < z < +\infty \tag{3.3-6}$$

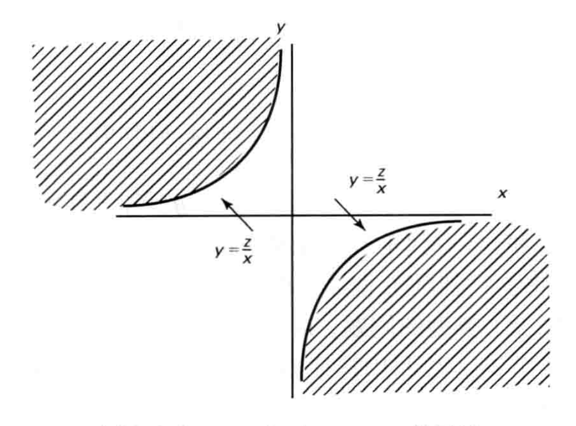


图3.3-4　$z<0$时，$xy \leqslant z$的区域

对于X和Y是独立的这种特例，即独立同分布(i. i. d.)的随机变量

$$f_X(x) = f_Y(x) \triangleq \frac{\alpha/\pi}{\alpha^2 + x^2}$$

① 注原文在这里是式(3.3-7)，疑有误——译者注。
② 原文在这里是式(3.3-7)，疑有误——译者注。

这就是所谓的柯西①概率准则。根据独立性可得

$$f_{XY}(x,y)=f_X(x)f_Y(y)$$

同时根据式(3.3-7)被积函数的对称性，通过变量替换可得②

$$\begin{aligned}f_Z(z)&=\left(\frac{\alpha}{\pi}\right)^2\int_0^{\infty}\frac{1}{z^2+\alpha^2x}\cdot\frac{1}{\alpha^2+x}\mathrm{d}x\\&=\left(\frac{\alpha}{\pi}\right)^2\frac{1}{z^2-\alpha^4}\ln\frac{z^2}{\alpha^4}\end{aligned}\tag{3.3-7}$$

图 3.3-5 给出了 $\alpha=1$ 时的 $f_Z(z)$。

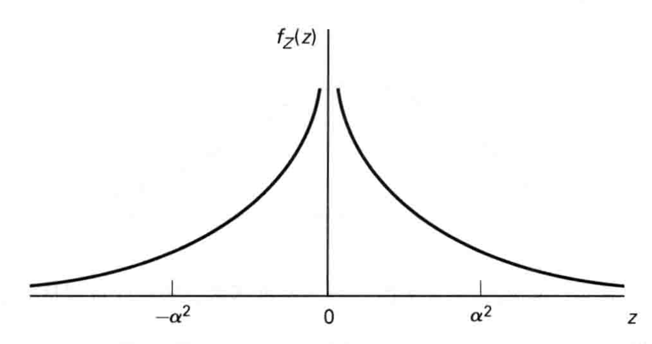


图 3.3-5　$Z=XY$ 的概率密度函数 $f_Z(z)$，X 和 Y 为独立同分布随机变量且满足柯西定律

例 3.3-2　(取最大值操作)如果 X 和 Y 是独立的随机变量，我们希望求 $Z=\max(X,Y)$ 的概率密度函数，即

$$F_Z(z)=P[\max(X,Y)\leqslant z]$$

由于事件 $\{\max(X,Y)\leqslant z\}$ 与事件 $\{X\leqslant z,Y\leqslant z\}$ 是等价的，因此

$$P[Z\leqslant z]=P[X\leqslant z,Y\leqslant z]=F_X(z)F_Y(z)\tag{3.3-8}$$

对两边求导可得

$$f_Z(z)=f_Y(z)F_X(z)+f_X(z)F_Y(z)\tag{3.3-9}$$

讨论一个特例，令 $f_X(x)=f_Y(x)$ 为区间 $[0,1]$ 上的均匀分布，可得

$$f_Z(z)=2z[u(z)-u(z-1)]\tag{3.3-10}$$

如图 3.3-6 所示。对于 $Z=\min(X,Y)$ 的概率密度函数，留做本章的习题。

例 3.3-3　(取指数函数的最大值)令 X 和 Y 为独立同分布的随机变量，概率密度函数 $f_X(x)=\mathrm{e}^{-x}u(x)$。设 $Z=\max(X,Y)$，试求 $f_Z(z)$ 以及概率值 $P[Z\leqslant 1]$。

解　根据 $P[Z\leqslant z]=P[X\leqslant z,Y\leqslant z]=P[X\leqslant z]P[Y\leqslant z]$ 可得

$$F_Z(z)=F_X(z)F_Y(z)=(1-\mathrm{e}^{-z})^2u(z)$$

以及

$$f_Z(z)=\frac{\mathrm{d}F_Z(z)}{\mathrm{d}z}=2\mathrm{e}^{-z}(1-\mathrm{e}^{-z})u(z)$$

概率密度函数如图 3.3-7 所示。最后可得，$F_Z(1)=(1-\mathrm{e}^{-1})^2u(1)\approx 0.4$。

① Auguste Louis Cauchy(1789－1857)，法国数学家，在天文学，光学，流体力学，函数论等方面著作颇丰。

② 参见 B. O. Pierce and R. M. Foster, *A Short Table of Integrals*, 4th ed. (Boston, MA: Ginn & Company, 1956), p. 8.

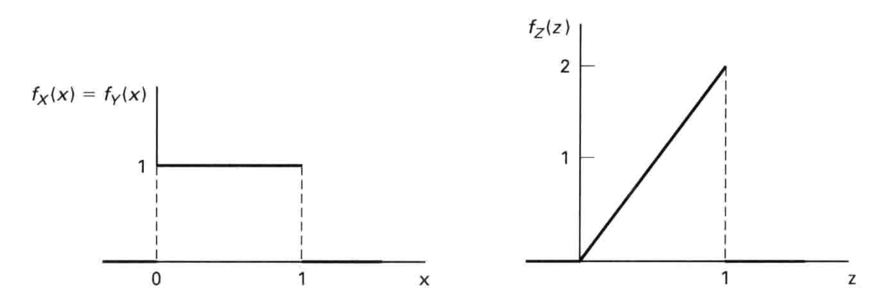


图3.3-6 $Z = \max(X, Y)$ 的概率密度函数，X 和 Y 独立同分布，在区间[0, 1]上均匀分布

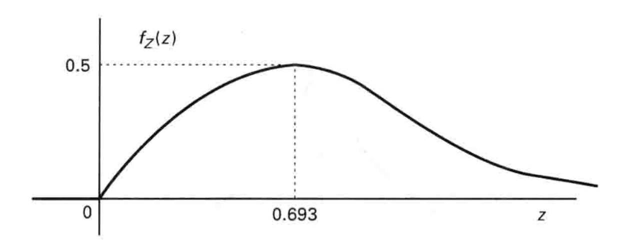


图3.3-7 取两个随机变量最大值的概率密度函数，随机变量服从指数分布

两个独立随机变量之和 在工程和科学问题中经常会遇到 $Z = X + Y$ 的情况(以及它的拓展 $Z = \sum_{i=1}^{N} X_i$)，所以此时的 $f_Z(z)$ 可能是 $Z = g(X, Y)$ 类型中最重要的概率密度函数。

类似于与其他类型的求解，我们需要找出满足 $\{Z \leqslant z\}$ 的点集 C_z，根据定义，此时事件 $\{X + Y \leqslant z\}$ 与事件 $\{(X, Y) \in C_z\}$ 是等价的。点集 C_z 是满足 $g(x, y) \triangleq x + y \leqslant z$ 中所有点的集合，即图3.3-8中直线左边的阴影区域。该阴影区域中的所有点都满足 $x + y \leqslant z$。

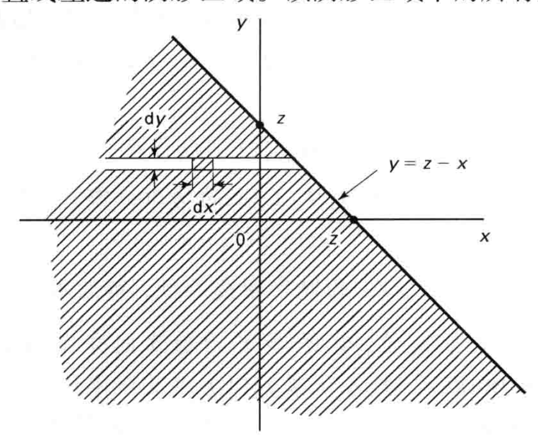


图3.3-8 用于计算 $Z \triangleq X + Y$ 概率密度函数的 C_z 区域(阴影部分)

根据式(3.3-2)可得

$$
\begin{aligned}
F_Z(z) &= \iint_{x+y\leqslant z} f_{XY}(x,y)\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-y} f_{XY}(x,y)\mathrm{d}x\right)\mathrm{d}y \qquad (3.3\text{-}11) \\
&= \int_{-\infty}^{\infty}[G_{XY}(z-y,y) - G_{XY}(-\infty,y)]\mathrm{d}y
\end{aligned}
$$

其中 $G_{XY}(x, y)$ 是不定积分，定义如下：

$$G_{XY}(x,y) \triangleq \int f_{XY}(x,y)\mathrm{d}x \tag{3.3-12}$$

通过对 $F_Z(z)$ 求导，可得概率密度函数表达式如下：

$$\begin{aligned} f_Z(z) = \frac{\mathrm{d}F_Z(z)}{\mathrm{d}z} &= \int_{-\infty}^{\infty} \frac{\mathrm{d}}{\mathrm{d}z}[G_{XY}(z-y,y)]\mathrm{d}y \\ &= \int_{-\infty}^{\infty} f_{XY}(z-y,y)\mathrm{d}y \end{aligned} \tag{3.3-13}$$

式(3.3-13)是一个非常重要的结论[与 $Z = XY$ 的式(3.3-6)相比较]。在许多时候，X 和 Y 是独立的随机变量，所以有 $f_{XY}(x, y) = f_X(x)f_Y(y)$。此时式(3.3-13)成为

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)\mathrm{d}y \tag{3.3-14}$$

这就是卷积积分，即 f_X 卷积 f_Y ①。利用变量替换 $x = z - y$，可很容易的证明式(3.3-14)可以写成如下表达式：

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\mathrm{d}x \tag{3.3-15}$$

例 3.3-4　(随机变量之和)令 X 和 Y 为独立的随机变量，且 $f_X(x) = \mathrm{e}^{-x}u(x)$，$f_Y(y) = \frac{1}{2}[u(y+1) - u(y-1)]$，令 $Z \triangleq X + Y$。试求 Z 的概率密度函数。

解　利用图形来跟踪函数轨迹是求解卷积类型问题的一个很有用的方法。$f_X(x)$ 和 $f_Y(y)$ 如图 3.3-9(a)所示，$f_X(z-y)$ 如图 3.3-9(b)所示。注意到 $f_X(z-y)$ 其实就是 $f_X(x)$ 翻转和平移变换后的结果。我们是如何知道翻转/平移变换后的图像最边缘的那个点就是 $y = z$ 呢？可考虑如下的等式：

$$f_X(z-y) = \mathrm{e}^{-(z-y)}u(z-y)$$

当 $y > z$ 时 $u(z-y) = 0$，因此翻转/平移后的方程只有在区间 $(-\infty, z]$ 上才不为 0，所以 $f_X(z-y)$ 最边缘的那个点就位于 $y = z$ 处。

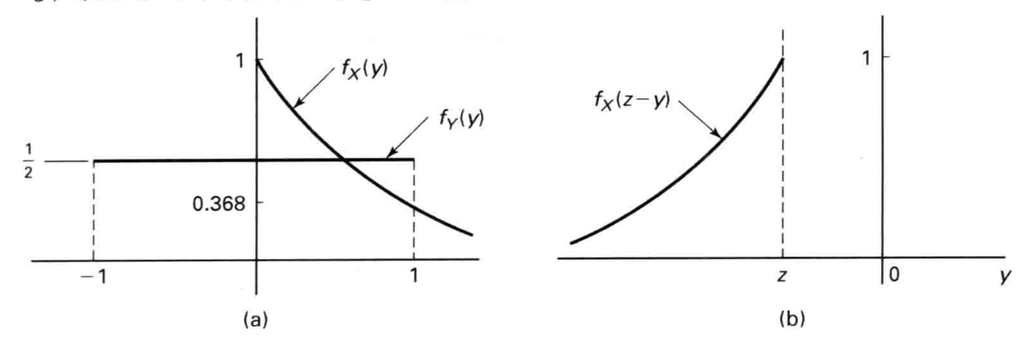


图 3.3-9　(a)概率密度函数 $f_X(y)$ 和 $f_Y(y)$；(b)翻转和平移后 $f_X(z-y)$ 的概率密度函数

① 常用 $f_Z = f_X * f_Y$ 来表示式(3.3-15)中的卷积积分。

由于f_X和f_Y都是非连续函数，因此我们不要期望对于所有的z都有相同的$f_Z(z)$表达式。这意味着我们必须必须根据z的不同取值，详细讨论式(3.3-14)。

(a) 区间1，$z<-1$。图3.3-10(a)给出了$z<-1$的情况。由于在此区间二者没有重叠部分，因此式(3.3-14)为0，也是$z<-1$时$f_Z(z)=0$。

(b) 区间2，$-1\leqslant z<1$。此区间的情况如图3.3-10(b)所示。式(3.3-14)可写为

$$f_Z(z)=\frac{1}{2}\int_{-1}^{z}\mathrm{e}^{-(z-y)}\mathrm{d}y$$
$$=\frac{1}{2}[1-\mathrm{e}^{-(z+1)}]$$

(c) 区间3，$z\geqslant 1$。这个区间的情况如图3.3-10(c)所示。利用式(3.3-14)可得

$$f_Z(z)=\frac{1}{2}\int_{-1}^{1}\mathrm{e}^{-(z-y)}\mathrm{d}y$$
$$=\frac{1}{2}[\mathrm{e}^{-(z-1)}-\mathrm{e}^{-(z+1)}]$$

如果要将得到的结果综合在一个图形上，必须先进行一个最终的重要观测：由于在计算中并不涉及任何的δ函数，因此$f_Z(z)$应该是z的一个连续函数。因此，作为对结果的一个检验，在区间的交界处$f_Z(z)$的结果应该吻合才对。例如，对于区间2和区间3的交界处，即$z=1$时，分析下式：

$$\tfrac{1}{2}[1-\mathrm{e}^{-(z+1)}]_{z=1}\overset{?}{=}\tfrac{1}{2}[\mathrm{e}^{-(z-1)}-\mathrm{e}^{-(z+1)}]_{z=1}$$

很显然，等式左右两端是相等的，因此我们对得到的结果应该有些自信(参见图3.3-11)。

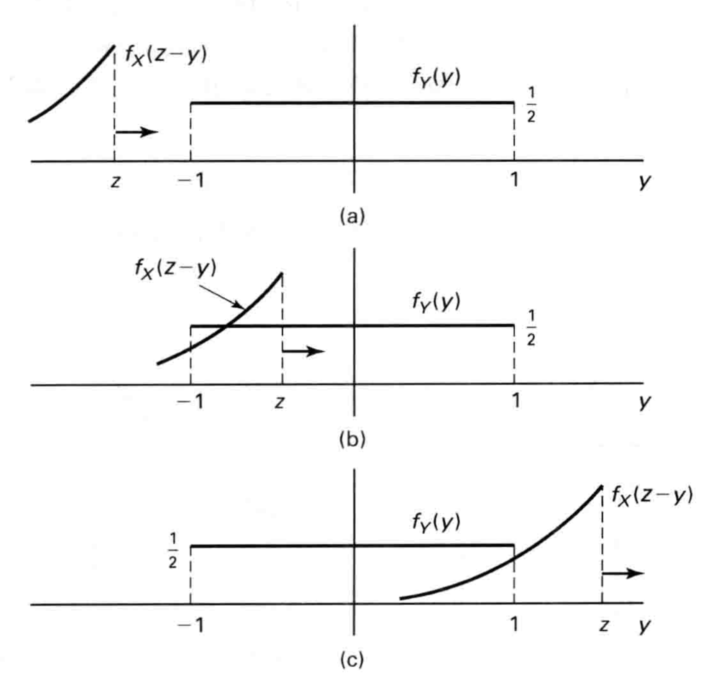


图3.3-10 $f_X(z-y)$与$f_Y(y)$的相互位置关系。(a) $z<1$; (b) $-1\leqslant z<1$; (c) $z>1$

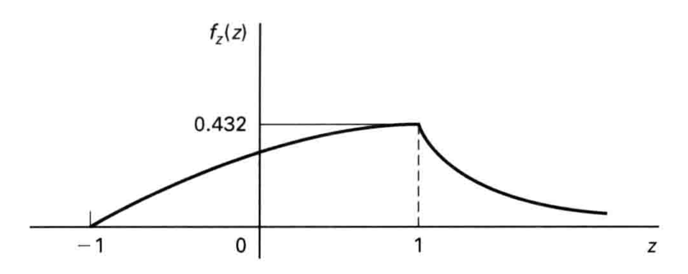


图 3.3-11 例 3.3-4 中的概率密度函数 $f_Z(z)$

可以很容易的将式(3.3-14)和式(3.3.15)拓展到 $Z = aX + bY$ 的情况。不失一般性，令 $a > 0$, $b > 0$。那么 $g(x, y) \triangleq ax + by \leqslant z$ 就表示直线 $y = z/b - ax/b$ 左边的区域(参见图 3.3-12)。

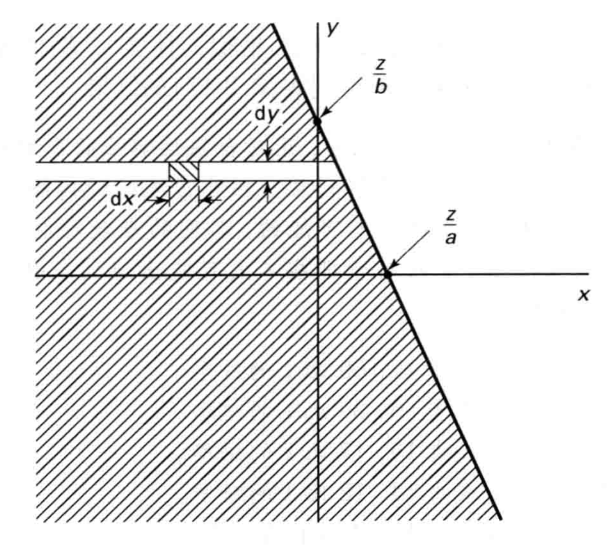


图 3.3-12 用于计算 $Z = aX + bY$ 概率密度函数积分的区域, $a > 0$, $b > 0$

因此有

$$\begin{aligned} F_Z(z) &= \iint_{g(x,y)\leqslant z} f_{XY}(x,y)\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} f_Y(y)\left(\int_{-\infty}^{z/a-by/a} f_X(x)\mathrm{d}x\right)\mathrm{d}y \end{aligned} \tag{3.3-16}$$

类似地，我们仍然通过对 z 求导来计算 $f_Z(z)$

$$f_Z(z) = \frac{1}{a}\int_{-\infty}^{\infty} f_X\left(\frac{z}{a} - \frac{by}{a}\right) f_Y(y)\mathrm{d}y \tag{3.3-17}$$

在此假设 X 和 Y 是独立的随机变量。同样地，可以通过下面的方式来计算 $f_Z(z)$

$$V \triangleq aX$$

$$W \triangleq bY$$

$$Z \triangleq V + W$$

在此仍然假设 $a > 0$, $b > 0$, 以及 X 和 Y 是独立的。通过式(3.3-14)可得

$$f_Z(z) = \int_{-\infty}^{\infty} f_V(z-w) f_W(w) \mathrm{d}w$$

其中，通过式(3.2-2)可得

$$f_V(v) = \frac{1}{a} f_X\left(\frac{v}{a}\right)$$

以及

$$f_W(w) = \frac{1}{b} f_Y\left(\frac{w}{b}\right)$$

综上所得

$$f_Z(z) = \frac{1}{ab} \int_{-\infty}^{\infty} f_X\left(\frac{z-w}{a}\right) f_Y\left(\frac{w}{b}\right) \mathrm{d}w \tag{3.3-18}$$

尽管式(3. 3-18)和式(3. 3. 17)“看起来”不太相似，其实二者是等价的。我们只需把式(3.3-18)中的变量替换为 $y \triangleq w/b$ 就能得到式(3.3-17)的结果。

例 3.3-5 (掷骰子游戏)在美国许多地方，用对合法博彩业收取的税费来补贴公共教育，修路等。博彩场所通过对赔率的设定来获取利益。在非常流行的掷骰子游戏中，玩家需要猜三个同时掷出的骰子点数。每个点数都是有可能的，不同的点数对应不同的赔率。出现可能性很大的点数对应的收益往往很小，而对于那些不太可能出现的点数就设置了较大的赔率。在一个大型博彩场所，赔率的设置如下：

1. 三个骰子点数之和等于 4 或者 17，60 比 1 的赔率。
2. 三个骰子点数之和等于 5 或者 16，30 比 1 的赔率。
3. 三个骰子点数之和等于 6 或者 15，17 比 1 的赔率。
4. 三个骰子点数之和等于 7 或者 14，12 比 1 的赔率。
5. 三个骰子点数之和等于 8 或者 13，8 比 1 的赔率。
6. 三个骰子点数之和等于 9，10，11 或者 12，6 比 1 的赔率。

在此解释一下，60 比 1 的赔率是指：玩家押 1 美元，如果这个事件发生了，那么将得到 60 -1 美元(1 美元是本金)。那么计算各个事件发生的概率就非常有意思了。

解 所有可能的点数其实由三个独立同分布随机变量之和决定。令 X，Y 和 Z 分别表示三个骰子的点数。为了求出计算结果，我们需要用到两个连续卷积。因此对于两个骰子点数之和，$X+Y$ 的累积分布函数为 $P_{X+Y}(l) \triangleq \sum_{i=1}^{6} P_X(l-i)P_Y(i)$，结果如图 3.3-13 所示。为了计算三个随机变量之和的 $P_{X+Y+Z}(n)$，需要进行第二次卷积 $P_{X+Y+Z}(n) = \sum_{i=2}^{12} P_Z(n-i)P_{X+Y}(i)$。结果如图 3.3-14 所示。经过第二次卷积，我们得到了最终需要的结果。

在此定义一个所谓的“公平赔率”(FP)，意思是平均而言博彩场所既不会亏本，也不会从玩家那里赢钱[①]。如果玩家用 1 美元对事件 E 押注，对于公平赔率而言获得的收益应该是 0，有 $0 = -\$1.00 + \mathrm{FP} \times P[E]$，那么有 $\mathrm{FP} = 1/P[E]$。

① 很显然，博彩场所是需要通过大量的盈利来维持各种开支，比如薪水，日常水电开支等。这里“公平赔率”的定义忽略了这些开支。另外，何谓“平均”，我们将在第 4 章详细讨论。

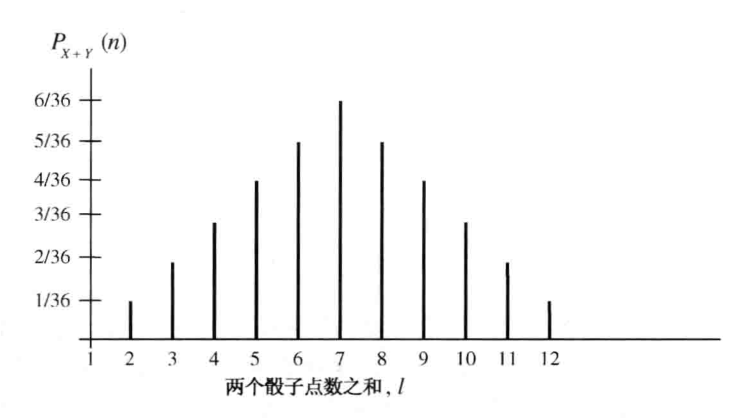


图 3.3-13 两个骰子点数之和的概率分布

通过图 3.3-14，我们可以直接给出如下的结果：

1. 点数之和为 4 或 17 的可能性为 3/216，那么公平赔率应该是 72∶1(与 60∶1 相比较)。
2. 点数之和为 5 或 16 的可能性为 6/216，那么公平赔率应该是 36∶1(与 30∶1 相比较)。
3. 点数之和为 6 或 15 的可能性为 10/216，那么公平赔率应该是 22∶1(与 17∶1 相比较)。
4. 点数之和为 7 或 14 的可能性为 15/216，那么公平赔率应该是 14∶1(与 12∶1 相比较)。
5. 点数之和为 8 或 13 的可能性为 21/216，那么公平赔率应该是 10∶1(与 8∶1 相比较)。
6. 点数之和为 9 或 12 的可能性为 25/216，那么公平赔率应该是 9∶1(与 6∶1 相比较)。
7. 点数之和为 10 或 11 的可能性为 27/216，那么公平赔率应该是 8∶1(与 6∶1 相比较)。

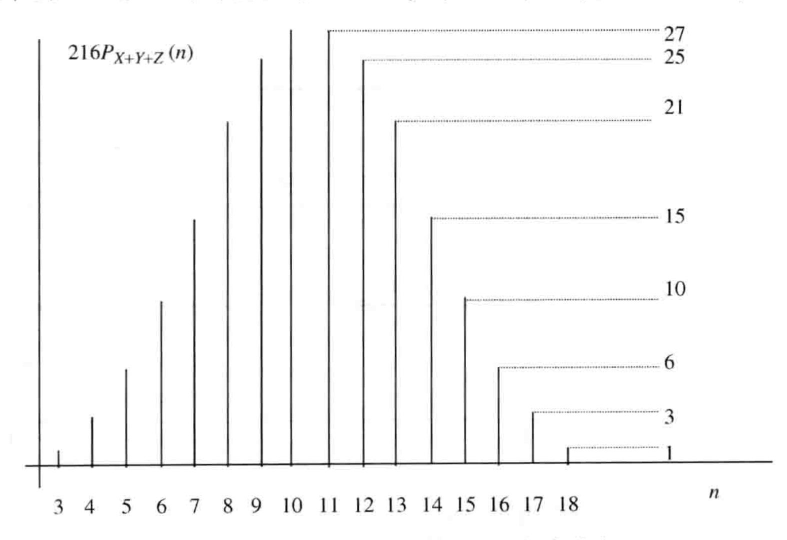


图 3.3-14 三个骰子点数之和的概率分布

例 3.3-6 (平方律检波器)令 X 和 Y 为独立同分布的随机变量，分布为 $U(-1,1)$。计算 $V \triangleq (X+Y)^2$ 的概率密度函数。

解 分两个步骤来求解这个问题。首先计算 $Z \triangleq X+Y$ 的概率密度函数，然后再计算 $V = Z^2$ 的概率密度函数。利用脉宽矩形函数可得

$$f_X(x) = \frac{1}{2}\text{rect}\left(\frac{x}{2}\right)$$

$$f_Y(y) = \frac{1}{2}\text{rect}\left(\frac{y}{2}\right)$$

$$f_X(z-y) = \frac{1}{2}\text{rect}\left(\frac{z-y}{2}\right)$$

根据式(3.3-14)可得

$$f_Z(z) = \frac{1}{4}\int_{-\infty}^{\infty} \text{rect}\left(\frac{y}{2}\right)\text{rect}\left(\frac{z-y}{2}\right)\text{d}y \tag{3.3-19}$$

计算式(3.3-19)最好的方法就是在图形上跟踪它的"移动",也就是跟z有关的方程$\text{rect}((z-y)/2)$与跟z无关的方程$\text{rect}(y/2)$相比较。这里所谓的"移动",是指随着z的变化,方程$f_X((z-y)/2)$好像在方程$f_Y(y)$上移动。z的4种取值情况如图3.3-15所示。

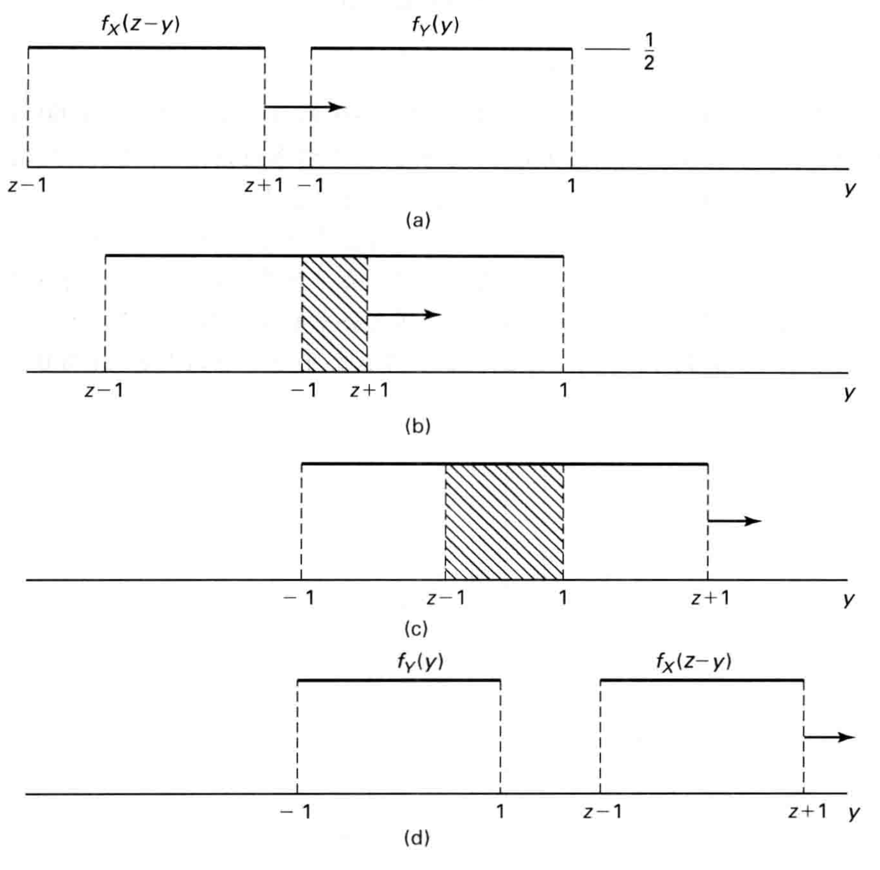


图3.3-15 在4个区间上两个均匀分布的卷积。(a) $z<-2$;(b) $-2\leqslant z<0$;(c) $0\leqslant z<2$;(d) $z\geqslant 2$

$f_Z(z)$在4个区间的计算如下:

(a) $z<-2$,该区域没有重合的区间,所以

$$f_Z(z) = 0$$

(b) $-2\leqslant z<0$,在该区域重合的区间为$(-1, z+1)$,所以

$$f_Z(z) = \frac{1}{4}\int_{-1}^{z+1}\text{d}y = \frac{1}{4}(z+2)$$

(c) $0 \leqslant z < 2$，在该区域重合的区间为 $(z-1, 1)$，所以

$$f_Z(z) = \frac{1}{4}\int_{z-1}^{1} \mathrm{d}y = \frac{1}{4}(2-z)$$

(d) $z \geqslant 2$，该区域没有重合的区间，所以

$$f_Z(z) = 0$$

综上所述，最终的结果为

$$f_Z(z) = \frac{1}{4}(2-|z|)\mathrm{rect}\left(\frac{z}{4}\right) \tag{3.3-20}$$

结果如图 3.3-16 所示。

为得到最终的解答，我们还需要计算 $V = Z^2$ 的概率密度函数。利用式(3.3-19)来计算 $f_V(v)$，其中 $g(z) = z^2$。当 $v > 0$ 时，方程 $v - z^2 = 0$ 有两个实根，分别是 $z_1 = \sqrt{v}$ 和 $z_2 = -\sqrt{v}$；当 $v < 0$ 时，方程没有实根。将式(3.3-20)代入下式

$$f_V(v) = \sum_{i=1}^{2} f_Z(z_i)/(2|z|)$$

可得

$$f_V(v) = \begin{cases} \frac{1}{4}\left(\frac{2}{\sqrt{v}} - 1\right) & 0 < v \leqslant 4 \\ 0, & \text{其他} \end{cases} \tag{3.3-21}$$

结果如图 3.3-17 所示。

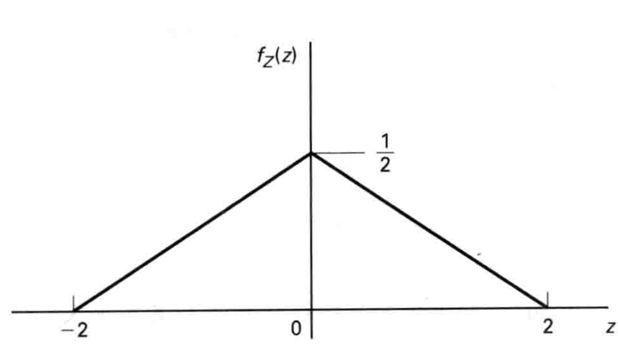


图 3.3-16　$Z = X + Y$ 的概率密度函数，X 和 Y 是独立同分布的随机变量，服从 $(-1, +1)$ 区间上的均匀分布

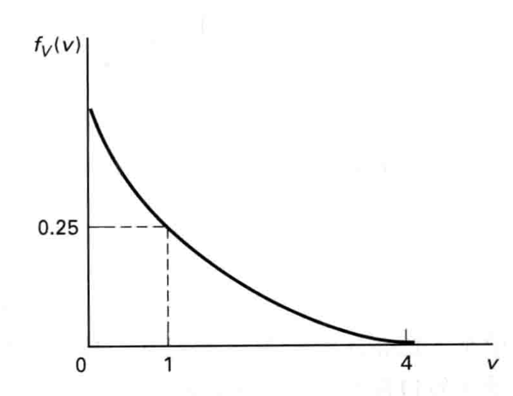


图 3.3-17　例 3.3-6 中 V 的概率密度函数

可以利用离散卷积来计算离散随机变量之和的概率密度函数。例如，设 X 和 Y 是两个随机变量，可能的取值分别是 $x_1, \cdots, x_k, \cdots$ 和 $y_1, \cdots, y_j, \cdots$。那么 $Z \triangleq X + Y$ 很显然也是离散的，它的概率质量函数为

$$P_Z(z_n) = \sum_{x_k + y_j = z_n} P_{X,Y}(x_k, y_j) \tag{3.3-22}$$

如果 X 和 Y 是独立的，式(3.3-22)可写为

$$P_Z(z_n) = \sum_{x_k + y_j = z_n} P_X(x_k)P_Y(y_j) = \sum_{x_k} P_X(x_k)P_Y(z_n - x_k) \tag{3.3-23a}$$

如果 z_n 和 x_k 是等距的①，那么式(3.3-22a)就可以看成一个离散卷积，又可以写为

$$P_Z(n)=\sum_{\text{所有}k}P_X(k)P_Y(n-k) \tag{3.3-23b}$$

在后面我们会学习如何使用式(3.3-22b)。

例 3.3-7 (伯努利分布随机变量之和)设 B_1 和 B_2 为两个独立的随机变量，二者服从伯努利分布，概率质量函数如下：

$$P_B(k)=\begin{cases}p,\ k=1\\ q,\ k=0, \quad q=1-p\\ 0,\ \text{其他}\end{cases}$$

令 $M\triangleq B_1+B_2$，求它的质量密度函数 $P_M(m)$。我们从一般性的结论入手

$$P_M(m)=\sum_{b=-\infty}^{+\infty}P_{B_1}(k)P_{B_2}(m-k)$$
$$=\sum_{b=0}^{1}P_{B_1}(k)P_{B_2}(m-k)$$

由于每个 B_i 的可能取值只能是 0 和 1，因此 M 可能的取值只能是 0，1 和 2。对于 m 取其他值的情况，$P_M(m)=0$。这个结果从图 3.3-18 的离散卷积中也可以看出来。

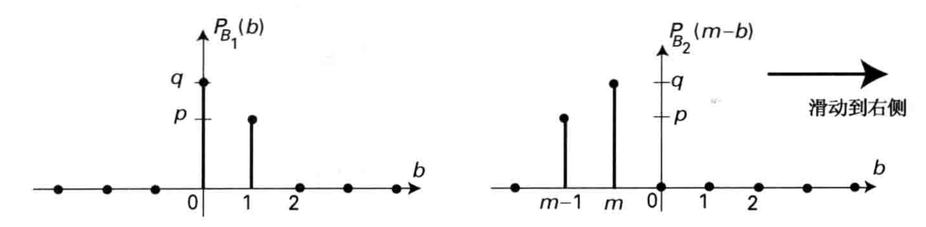


图 3.3-18　两个伯努利分布随机变量的离散卷积

现计算 P_M 中的非 0 值，可得

$$P_M(0)=P_{B_1}(0)P_{B_2}(0)=q^2$$
$$P_M(1)=P_{B_1}(0)P_{B_2}(1)+P_{B_1}(1)P_{B_2}(0)=2pq$$
$$P_M(2)=P_{B_1}(1)P_{B_2}(1)=p^2$$

学生可能会注意到 M 其实服从二项分布 $b(k;2,p)$。为什么会是这样的呢？如果再加入一个独立的伯努利随机变量，会发生什么情况呢？

例 3.3-8 (泊松分布随机变量之和)设 X 和 Y 是两个独立的随机变量，二者服从泊松分布，概率质量函数分布为 $P_X(k)=\frac{1}{k!}e^{-a}a^k$ 和 $P_Y(i)=\frac{1}{i!}e^{-b}b^i$，其中 a 和 b 分别是 X 和 Y 泊松分布的参数。令 $Z\triangleq X+Y$，那么 Z 的概率质量函数 $P_Z(n)$ 如下可得：

$$\begin{aligned}P_Z(n)&=\sum_{k=0}^{n}P_X(k)P_Y(n-k)\\&=\sum_{k=0}^{n}\frac{1}{k!}\frac{1}{(n-k)!}e^{-(a+b)}a^kb^{n-k}\end{aligned} \tag{3.3-24}$$

① *例如，$z_n=n\Delta$，$x_k=k\Delta$，其中 Δ 是一个常数。

根据二项式定理

$$\sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}=(a+b)^n \tag{3.3-25}$$

可得

$$\begin{aligned}P_Z(n)&=\frac{1}{n!}\mathrm{e}^{-(a+b)}\sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}\\&=\frac{(a+b)^n}{n!}\mathrm{e}^{-(a+b)},\qquad n\geqslant 0\end{aligned} \tag{3.3-26}$$

结果服从参数为 $a+b$ 的泊松分布。那么可得到如下重要的结论：参数分别为 a 和 b 的独立泊松随机变量，二者之和服从参数为 $(a+b)$ 的泊松分布。

例 3.3-9　（二项分布随机变量之和）现在再来求解一个比前一个例题更富挑战性的问题，即二项分布随机变量 X 和 Y 之和。令 $Z=X+Y$，那么概率质量函数 $P_Z(m)$ 如下可得：

$$P_Z(m)=\sum_{k=-\infty}^{\infty}P_X(k)P_Y(m-k)$$

其中

$$P_X(k)=P_Y(k)=\begin{cases}0, & k<0\\ \binom{n}{k}p^k q^{n-k}, & 0\leqslant k\leqslant n\\ 0, & k>n\end{cases}$$

那么可得

$$\begin{aligned}P_Z(m)&=\sum_{k=\max(0,m-n)}^{\min(n,m)}\binom{n}{k}p^k q^{n-k}\binom{n}{m-k}p^{m-k}q^{n-(m-k)}\\&=p^m q^{2n-m}\sum_{k=\max(0,m-n)}^{\min(n,m)}\binom{n}{k}\binom{n}{m-k}\end{aligned}$$

上面的求和公式需要限定 $0\leqslant k\leqslant n$ 和 $0\leqslant m-k\leqslant n$，其中后一个条件与 $m-n\leqslant k\leqslant m$ 等价。还可知求和的取值范围应该是 $\max(0,\ m-n)\leqslant k\leqslant\min(n,\ m)$。

再利用一个神奇的结论

$$\sum_{k=\max(0,m-n)}^{\min(n,m)}\binom{n}{k}\binom{n}{m-k}=\binom{2n}{m} \tag{3.3-27}$$

可得到 Z 的概率质量函数如下：

$$P_Z(m)=\binom{2n}{m}p^m q^{2n-m}\triangleq b(m;2n,p) \tag{3.3-28}$$

综上可得，两个独立同分布的二项分布随机变量之和仍然服从二项分布，其中求和之前概率质量函数为 $b(k;\ n,\ p)$，求和后的概率质量函数为 $b(k;\ 2n,\ p)$。

为了验证式(3.3-27)成立与否，我们注意到无论是 $m>n$（即从 $k=m-n$ 到 $k=n$ 求和）还是 $m\leqslant n$（即从 $k=0$ 到 $k=m$ 求和），左边表达式的值都相等。一个简单的方法就是把

左边的这个表达式在两种情况下分别展开。当 $m \leqslant n$ 时，展开情况如下：

$$\binom{n}{0}\binom{n}{m}+\binom{n}{1}\binom{n}{m-1}+\dots+\binom{n}{m}\binom{n}{0} \tag{3.3-29}$$

当 $m > n$ 时，展开可以得到同样的结果。

进一步验证式(3.3-27)，注意到子总体 m 的取值由总体 $2n$ 决定，为 C_m^{2n}。但是也可以由另外一种方法来计算子总体的数值，可将总体 $2n$ 拆成两个分别是 n 的总体。设这两个总体分别叫做 A 和 B。那么乘积 $C_k^n C_{m-k}^n$ 可以看成从 A 中 k 个子总体到 B 中 $m-k$ 个子总体所有可能的路线。这样就有

$$\sum_{k=0}^{m} C_k^n C_{m-k}^n = C_m^{2n} \tag{3.3-30}$$

根据前面提及的

$$\sum_{k=m-n}^{n} C_k^n C_{m-k}^n = \sum_{k=0}^{m} C_k^n C_{m-k}^n \text{①}$$

可知当 k 从 $m-n$ 到 n 变化时，式(3.3-27)是成立的。

在第 4 章，我们还会学习一种简单的方法来证明独立同分布的二项分布随机变量之和仍然服从二项分布。这种方法用到的变换称为矩生成函数以及(或者)特征函数。

我们曾在 3.2 节提到过，尽管式(3.3-23a)(该公式的拓展将在 3.4 节介绍)用来求解这种类型的问题非常方便，但非直接法有的时候会更简单。接下来我们用一些例题来介绍。

例 3.3-10　(平方和)令 X 和 Y 为独立同分布的高斯随机变量，服从 $X:N(0, \sigma^2)$。求 $Z \triangleq X^2+Y^2$ 的概率密度函数。

解　根据式(3.3-2)中的基本结论

$$\begin{aligned} F_Z(z) &= \iint_{(x,y)\in C_z} f_{XY}(x,y)\mathrm{d}x\,\mathrm{d}y, \qquad z \geqslant 0 \\ &= \frac{1}{2\pi\sigma^2}\iint_{x^2+y^2\leqslant z} \mathrm{e}^{-(1/2\sigma^2)(x^2+y^2)}\mathrm{d}x\,\mathrm{d}y \end{aligned} \tag{3.3-31}$$

区域 C_z 即图 3.3-19 中的阴影区域。

在极坐标体系中更方便计算式(3.3-31)。令

$$x = r\cos\theta \qquad y = r\sin\theta$$

$$\mathrm{d}x\,\mathrm{d}y \to r\mathrm{d}r\mathrm{d}\theta$$

那么 $x^2+y^2 \leqslant z \to r \leqslant \sqrt{z}$，此时式(3.3-31)转变为

$$\begin{aligned} F_Z(z) &= \frac{1}{2\pi\sigma^2}\int_0^{2\pi}\mathrm{d}\theta\int_0^{\sqrt{z}} r\exp\left(-\frac{1}{2\sigma^2}r^2\right)\mathrm{d}r \\ &= [1-\mathrm{e}^{-z/2\sigma^2}]u(z) \end{aligned} \tag{3.3-32}$$

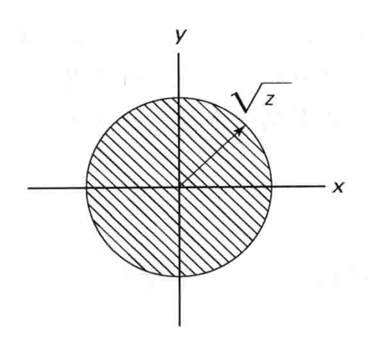


图 3.3-19　事件 $\{X^2+Y^2 \leqslant z\}$ 的区域 C_z，$z \geqslant 0$

---

① 也可以对右边表达式进行变量替换 $l \triangleq m-k$ 验证这个公式。此时的求和顺序是从大到小的，但翻转求和的顺序不会影响求和的结果。

以及

$$f_Z(z)=\frac{\mathrm{d}F_Z(z)}{\mathrm{d}z}=\frac{1}{2\sigma^2}\mathrm{e}^{-z/2\sigma^2}u(z) \tag{3.3-33}$$

综上，当 X 和 Y 是独立同分布的零均值高斯随机变量时，$Z=X^2+Y^2$ 是一个服从指数分布的随机变量。

例 3.3-11 （平方和开方）把前面一个例题修改为求 $Z\triangleq(X^2+Y^2)^{1/2}$ 的概率密度函数，得到的结果完全不同。同样地，我们仍然利用式(3.3-2)，只不过此时 C_z 为图 3.3-20 中的阴影区域。

因此有

$$\begin{aligned}F_Z(z)&=\frac{1}{2\pi\sigma^2}\int_0^{2\pi}\mathrm{d}\theta\int_0^z r\exp\left(-\frac{1}{2\sigma^2}r^2\right)\mathrm{d}r\\&=(1-\mathrm{e}^{-z^2/2\sigma^2})u(z)\end{aligned} \tag{3.3-34}$$

以及

$$f_Z(z)=\frac{z}{\sigma^2}\mathrm{e}^{-z^2/2\sigma^2}u(z) \tag{3.3-35}$$

上式为瑞利概率密度函数，这也可以看成自由度为 2 的 Chi 平方分布。图 3.3-21 同时给出了指数分布和瑞利分布的概率密度函数。

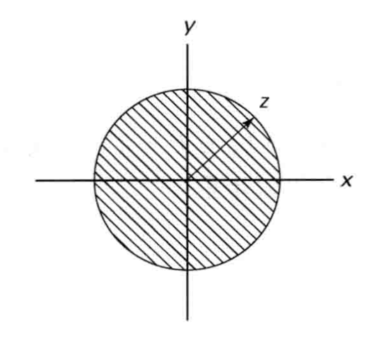


图 3.3-20 事件 $\{(X^2+Y^2)^{1/2}\leqslant z\}$ 的区域 C_z

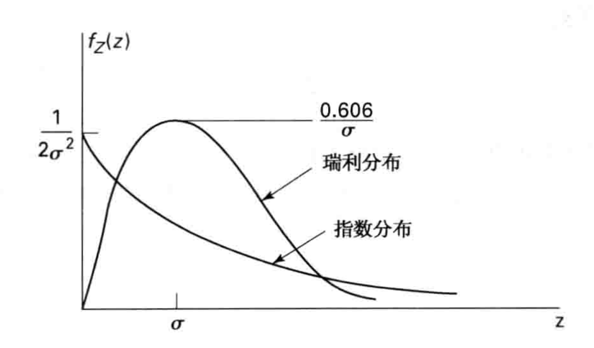


图 3.3-21 瑞利和指数分布的概率密度函数

早在 20 世纪 40 年代，Stephen O. Rice[3-3]就在电噪声方面做了诸多先驱性的工作，他的研究表明具有中心频率 ω 的窄带噪声信号可以用如下的模型表示：

$$Z(t)=X\cos\omega t+Y\sin\omega t \tag{3.3-36}$$

其中 t 是时间，ω 是弧度频率，单位为 rad/s，X 和 Y 是独立同分布的高斯随机变量，概率密度函数为 $N(0,\sigma^2)$。包络 $Z\triangleq(X^2+Y^2)^{1/2}$ 服从参数为 σ 的瑞利分布。

下一个例题将把例 3.3-10 的结论一般化，将得到通信理论中一个非常有用的结论。

***例 3.3-12**① （Rice 分布）②S. O. Rice 曾经考虑过这样一个问题：设 X 和 Y 是独立的高斯随机变量，且 X: $N(P,\sigma^2)$，Y: $N(0,\sigma^2)$，试求 $Z=(X^2+Y^2)^{1/2}$ 的概率密度函数。注意到当 $P=0$ 时就是例 3.3-11 的结论。

① 带 * 号的例题有些难度，在第一次学习的时候可先不考虑这些例题。

② 有的时候也称为 Rice-Nakagami 分布，这是出于尊重 Nakagami 在第二次世界大战期间对此做出的贡献。

可写出如下表达式：

$$F_Z(z)=\begin{cases}\dfrac{1}{2\pi\sigma^2}\displaystyle\iint_{(x^2+y^2)^{1/2}\leqslant z}\exp\left[-\frac{1}{2}\left(\left[\frac{x-P}{\sigma}\right]^2+\left(\frac{y}{\sigma}\right)^2\right)\right]\mathrm{d}x\,\mathrm{d}y, & z>0\\ 0, & z<0\end{cases}\tag{3.3-37}$$

根据笛卡儿坐标系到极坐标系的转换，即 $x = r\cos\theta$，$y = r\sin\theta$，$r = (x^2 + y^2)^{1/2}$，$\theta = \arctan(y/x)$，可得

$$F_Z(z)=\frac{\exp\left[-\frac{1}{2}\left(\frac{P}{\sigma}\right)^2\right]}{2\pi\sigma^2}\int_0^z \mathrm{e}^{-\frac{1}{2}(r/\sigma)^2}\left(\int_0^{2\pi}\mathrm{e}^{rP\cos\theta/\sigma^2}\mathrm{d}\theta\right)r\mathrm{d}r\cdot u(z)\tag{3.3-38}$$

其中下面的方程被称为第一类零阶修正贝塞尔方程，该方程单调递增类似于指数函数 e^x。

$$I_o(x)\triangleq\frac{1}{2\pi}\int_0^{2\pi}\mathrm{e}^{x\cos\theta}\mathrm{d}\theta$$

利用上面的记号，复杂的式(3.3-38)可简化为

$$F_Z(z)=\frac{\exp\left[-\frac{1}{2}\left(\frac{P}{\sigma}\right)^2\right]}{\sigma^2}\int_0^z rI_o\left(\frac{rP}{\sigma^2}\right)\mathrm{e}^{-\frac{1}{2}(r/\sigma)^2}\mathrm{d}r\cdot u(z)\tag{3.3-39}$$

其中阶跃函数 $u(z)$ 确保对于所有的 z 上式都成立。为了得到 $f_Z(z)$，我们对上式关于 z 求导得到

$$f_Z(z)=\frac{z}{\sigma^2}\exp\left[-\frac{1}{2}\left(\frac{P^2+z^2}{\sigma^2}\right)\right]I_o\left(\frac{zP}{\sigma^2}\right)\cdot u(z)\tag{3.3-40}$$

式(3.3-40)给出的表达式称为 Rice 概率密度函数。由于 $I_o(0) = 1$，当 $P = 0$ 时就成为瑞利分布。当 $zP \gg \sigma^2$，即 $I_o(\cdot)$ 的自变量很大时，我们采用如下的近似计算：

$$I_o(x)\approx\frac{\mathrm{e}^x}{(2\pi x)^{1/2}}$$

可得

$$f_Z(z)\approx\frac{1}{\sqrt{2\pi\sigma^2}}\left(\frac{z}{P}\right)^{1/2}\mathrm{e}^{-\frac{1}{2}[(z-P)/\sigma]^2}$$

此时基本上成为高斯分布[除了多了一个因子 $(z/P)^{1/2}$]。这就是一个较强的正弦函数与一个较弱的窄带高斯噪声之和的包络的概率密度函数，这种情况在电子通信中不太容易发生。

3.4　$V = g(X, Y)$，$W = h(X, Y)$ 型问题的求解

对于具有两个随机变量的两个方程，其实可以看成前面问题的一种拓展，只不过这个时候计算起来有些麻烦。

基础问题

已知两个随机变量 X, Y, 它们的联合概率密度函数 $f_{XY}(x, y)$，两个可微函数 $g(x, y)$ 和 $h(x, y)$。根据 $V = g(X, Y)$ 和 $W = h(X, Y)$ 来产生两个新的随机变量。试问如何计算 V 和 W 的联合累积分布函数 $F_{VW}(v, w)$ [或者联合概率密度函数 $f_{VW}(v, w)$]？

问题描述

1. 通信系统中经常遇到的一种变换关系如图 3.4-1 所示，例如立体声基带系统的参数[3-2]。其中 $\{a_i\}$ 表示增益。如果 $a_1 = a_2 = \cos\theta$，$a_3 = a_4 = \sin\theta$，那么这个电路就是一个典型的“θ 转动变换器”。在另外一种情况下，如果用 X 和 Y 来分别表示立体声播放系统中的左右通道采集来的信号，当所有的 a_i 都设定为 1 时，V 和 W 就分别表示差信号及和信号。此时就可以用和差信号来产生待传送的信号。如果假设没有源信号，X 和 Y 就只表示高斯噪声，那么 V 和 W 的概率密度函数是什么？
2. 飞机的着陆点同预定着陆位置之间的误差，在笛卡儿坐标体系中用(X, Y)来表示。我们希望在极坐标体系中描述这种误差，即 $V \triangleq (X^2+Y^2)^{1/2}$，$W \triangleq \arctan(Y/X)$。如果已知笛卡儿坐标体系中着陆误差的联合概率密度函数$f_{XY}(x, y)$，如何计算极坐标体系中着陆误差的概率密度函数？

求解该问题就是需要求出点集 C_{vw}，使得事件 $\{V \leqslant v, W \leqslant w\}$ 和事件 $\{(X, Y) \in C_{vw}\}$ 等价①。此时，基本联系为

$$\begin{aligned} P[V \leqslant v, W \leqslant w] &\triangleq F_{VW}(v,w) \\ &= \iint_{(x,y)\in C_{vw}} f_{XY}(x,y)\mathrm{d}x\,\mathrm{d}y \end{aligned} \tag{3.4-1}$$

区域 C_{vw} 由点 x 和 y 来设定，需满足下面的等式：

$$C_{vw} = \{(x,y)\colon g(x,y) \leqslant v, h(x,y) \leqslant w\} \tag{3.4-2}$$

下面我们将用一个例题来讲解式(3.4-1)的使用。

例 3.4-1　(和与差)已知 $V \triangleq X+Y$，$W \triangleq X-Y$，试求概率密度函数$f_{VW}(v, w)$。点集 C_{vw} 由条件 $g(x, y) \triangleq x + y \leqslant v$ 和 $h(x, y) \triangleq x - y \leqslant w$ 约束而成。当 $v > 0$，$w > 0$ 时如图 3.4-2 所示。

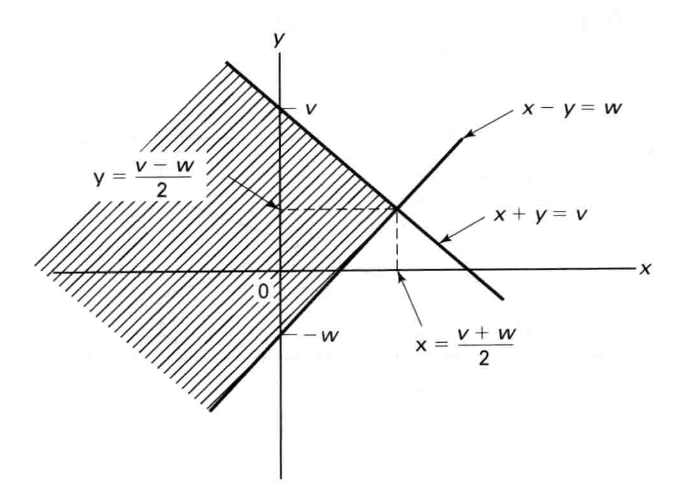


图 3.4-2　例 3.4-1 中的点集 C_{vw}（图中阴影部分）

在阴影区域的积分为

① 采用更加详细的记号，应该写为 $\{\zeta: V(\zeta) \leqslant v, W(\zeta) \leqslant w\} = \{\zeta: (X(\zeta), Y(\zeta)) \in C_{vw}\}$。

$$F_{VW}(v,w)=\int_{-\infty}^{(v+w)/2}\left(\int_{x-w}^{v-x}f_{XY}(x,y)\mathrm{d}y\right)\mathrm{d}x \tag{3.4-3}$$

为了计算联合概率密度函数$f_{VW}(v, w)$，在此利用式(2.6-30)的结论，可得

$$\begin{aligned}
f_{VW}(v,w)&=\frac{\partial^2 F_{VW}(v,w)}{\partial v\,\partial w}\\
&=\frac{\partial^2}{\partial v\,\partial w}\int_{-\infty}^{(v+w)/2}\left(\int_{x-w}^{v-x}f_{XY}(x,y)\mathrm{d}y\right)\mathrm{d}x\\
&=\frac{\partial}{\partial v}\left[\frac{\partial}{\partial w}\int_{-\infty}^{(v+w)/2}\left(\int_{x-w}^{v-x}f_{XY}(x,y)\mathrm{d}y\right)\mathrm{d}x\right]\\
&=\frac{\partial}{\partial v}\left[\frac{1}{2}\int_{(v-w)/2}^{(v-w)/2}f_{XY}\left(\frac{v+w}{2},y\right)\mathrm{d}y+\int_{-\infty}^{(v+w)/2}\left(\frac{\partial}{\partial w}\int_{x-w}^{v-x}f_{XY}(x,y)\mathrm{d}y\right)\mathrm{d}x\right]\\
&=\frac{\partial}{\partial v}\left[\int_{-\infty}^{(v+w)/2}f_{XY}(x,x-w)\mathrm{d}x\right]
\end{aligned}$$

因为对于连续随机变量X和Y第一个积分值为0，所以有

$$\begin{aligned}
&=\frac{\partial}{\partial v}\int_{-\infty}^{(v+w)/2}f_{XY}(x,x-w)\mathrm{d}x\\
&=\frac{1}{2}f_{XY}\left(\frac{v+w}{2},\frac{v-w}{2}\right)
\end{aligned} \tag{3.4-4}$$

在此利用了对积分求导的一般公式(参见附录A.2)。这个简单的问题涉及了线性方程和两个随机变量，需要谨慎求解才能得到联合概率密度函数的表达式。因此，对于本节这种类型的问题，以及扩展到n个随机变量的情况，即$Y_1=g_1(X_1,\cdots,X_n)$，$Y_2=g_2(X_1,\cdots,X_n)$，$\cdots$，$Y_n=g_n(X_1,\cdots,X_n)$，需要用到一种称为“直接法”的技术，该方法可用来求解联合概率密度函数。其实这是式(3.2-23)在二维情况下的一种扩展。

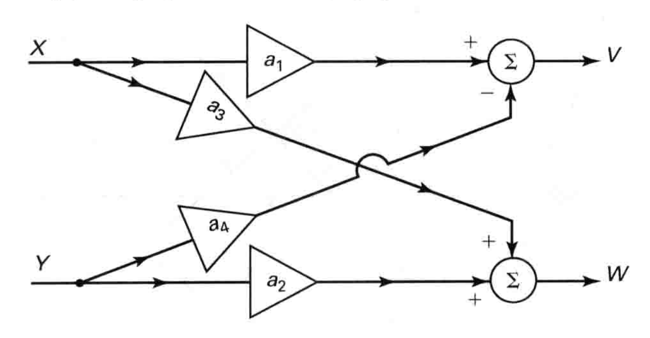


图3.4-1　一个用于二元变量之间转换的矩阵

由 f_{XY} 直接求 f_{VW}

我们在此不再采用式(3.4-1)，改用一种新的方法来求$f_{VW}(v, w)$。考虑如下的基本事件

$$\{v<V\leqslant v+\mathrm{d}v, w<W\leqslant w+\mathrm{d}w\}$$

以及一一映射[①]的可微方程 $v = g(x, y)$ 和 $w = h(x, y)$。逆映射存在，为 $x = \phi(v, w)$ 和 $y = \psi(v, w)$。后面将考虑更一般的情况，即一组以上的 (x_i, y_i) 来产生给定的 (v, w)。

概率 $P[v < V \leqslant v + \mathrm{d}v, w < W \leqslant w + \mathrm{d}w]$ 是指 V 和 W 落在一个无限小矩形区域内的概率，该矩形区域的四个顶点分别为 (v, w)，$(v + \mathrm{d}v, w)$，$(v, w + \mathrm{d}w)$ 和 $(v + \mathrm{d}v, w + \mathrm{d}w)$。这个矩形区域在 x，y 坐标系的像是[②]一个无限小的平行四边形，四个顶点分别是

$$P_1 = (x, y)$$

$$P_2 = \left(x + \frac{\partial \phi}{\partial v}\mathrm{d}v, y + \frac{\partial \psi}{\partial v}\mathrm{d}v\right)$$

$$P_3 = \left(x + \frac{\partial \phi}{\partial w}\mathrm{d}w, y + \frac{\partial \psi}{\partial w}\mathrm{d}w\right)$$

$$P_4 = \left(x + \frac{\partial \phi}{\partial v}\mathrm{d}v + \frac{\partial \phi}{\partial w}\mathrm{d}w, y + \frac{\partial \psi}{\partial v}\mathrm{d}v + \frac{\partial \psi}{\partial w}\mathrm{d}w\right)$$

映射关系如图 3.4-3 所示。

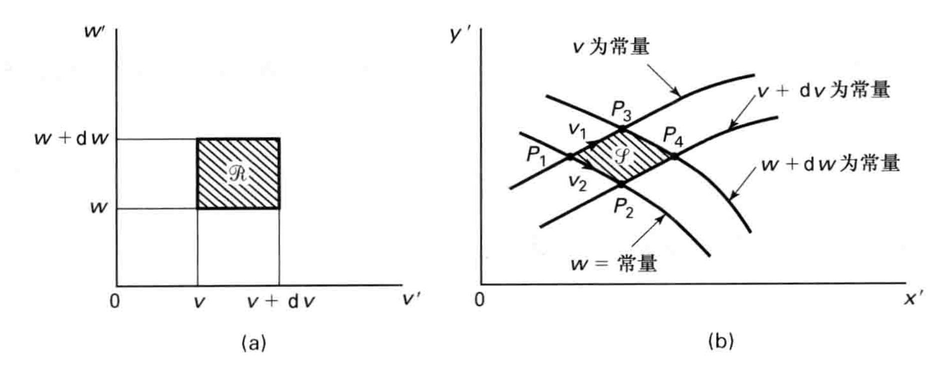


图 3.4-3　(a) v-w 坐标系中一个无限小的矩形区域；(b) 映射到 x-y 坐标系中一个无限小的平行四边形

令 $\mathscr{R}$ 表示图 3.4-3(a) 中的矩形区域，$\mathscr{S}$ 表示图 3.4-3(b) 中的平行四边形区域，且用 $A(\mathscr{R})$ 和 $A(\mathscr{S})$ 分别表示区域 $\mathscr{R}$ 和 $\mathscr{S}$ 的面积，那么可得

$$P[v < V \leqslant v + \mathrm{d}v, w < W \leqslant w + \mathrm{d}w] = \iint_{\mathscr{R}} f_{VW}(\zeta, \eta)\mathrm{d}\zeta\,\mathrm{d}\eta \tag{3.4-5}$$

$$= f_{VW}(v, w)A(\mathscr{R}) \tag{3.4-6}$$

$$= \iint_{\mathscr{S}} f_{XY}(\xi, \eta)\mathrm{d}\zeta\,\mathrm{d}\eta \tag{3.4-7}$$

$$= f_{XY}(x, y)A(\mathscr{S}) \tag{3.4-8}$$

其中，式(3.4-5)由式(3.4-1)中给出的基本关系得出；式(2.4-6)由式(3.4-6)中对概率密度函数的解释得出；式(3.4-7)由对点集 $\mathscr{S}$ 的定义得出，即 $\mathscr{S}$ 为使得事件 $\{(V, W) \in \mathscr{R}\}$ 和事件 $\{(X, Y) \in \mathscr{S}\}$ 等价的点的集合；式(3.4-8)再次用到了关于概率密度函数的解释。

通过式(3.4-6)和式(3.4-8)可得

---

① 每个点 (x, y) 只映射到唯一的点 (v, w)，反之亦然。

② 参见例[3-4, p.769]。

$$f_{VW}(v,w)=\frac{A(\mathscr{S})}{A(\mathscr{R})}f_{XY}(x,y) \tag{3.4-9}$$

其中 $x=\phi(v, w)$, $y=\psi(v, w)$。

至此可知，其实只需要计算两个面积之比。这部分内容已在附录 C 里面完成。在此我们将要认识到，面积之比 $A(\mathscr{S})/A(\mathscr{R})$ 是雅可比变换 $x=\phi(v, w)$, $y=\psi(v, w)$ 数值的大小，在此用 $\tilde{J}$ 表示。如果方程 $v=g(x, y)$, $w=h(x, y)$ 的解不止一个，即 $x_1=\phi_1(v, w)$, $y_1=\psi_1(v, w)$, $x_2=\phi_2(v, w)$, $y_2=\psi_2(v, w)$, $\cdots$, $x_n=\phi_n(v, w)$, $y_n=\psi_n(v, w)$，那么区域 $\mathscr{R}$ 映射到多个不相交的区域 $\mathscr{S}_1$, $\mathscr{S}_2$, $\cdots$, $\mathscr{S}_n$，以及 $A(\mathscr{S}_i)/A(\mathscr{R})=|\tilde{J}_i|$, $i=1, \cdots, n$。其中 $|\tilde{J}_i|$ 经常被写成行列式的形式

$$|\tilde{J}_i|=\text{mag}\begin{vmatrix}\partial\phi_i/\partial v & \partial\phi_i/\partial w\\ \partial\psi_i/\partial v & \partial\psi_i/\partial w\end{vmatrix}=|\partial\phi_i/\partial v\times\partial\psi_i/\partial w-\partial\psi_i/\partial v\times\partial\phi_i/\partial w| \tag{3.4-10}$$

最终结果得到如下重要的公式

$$f_{VW}(v,w)=\sum_{i=1}^{n}f_{XY}(x_i,y_i)|\tilde{J}_i| \tag{3.4-11}$$

根据附录 C 可知 $|\tilde{J}_i^{-1}|=|\tilde{J}_i|\triangleq|\partial g/\partial x_i\times\partial h/\partial y_i-\partial g/\partial y_i\times\partial h/\partial x_i|$，那么可得到该重要公式的等价形式

$$f_{VW}(v,w)=\sum_{i=1}^{n}f_{XY}(x_i,y_i)/|J_i| \tag{3.4-12}$$

例 3.4-2 (线性方程)已知如下两个方程：

$$\begin{aligned}v&\triangleq g(x,y)=3x+5y\\ w&\triangleq h(x,y)=x+2y\end{aligned} \tag{3.4-13}$$

以及这两个随机变量 X 和 Y 的联合概率密度函数 f_{XY}。试求两个新的随机变量 $V=g(X, Y)$ 和 $W=h(X, Y)$ 的联合概率密度函数。

解 根据式(3.4-13)求出逆映射

$$\begin{aligned}x&=\phi(v,w)=2v-5w\\ y&=\psi(v,w)=-v+3w\end{aligned}$$

可得

$$\frac{\partial\phi}{\partial v}=2,\frac{\partial\phi}{\partial w}=-5,\frac{\partial\psi}{\partial v}=-1,\frac{\partial\psi}{\partial w}=3$$

以及

$$|\tilde{J}|=\text{mag}\begin{vmatrix}2 & -5\\ -1 & 3\end{vmatrix}=1$$

在此假设 $f_{XY}(x, y)=(2\pi)^{-1}\exp[-\frac{1}{2}(x^2+y^2)]$。那么根据式(3.4-11)可得

$$\begin{aligned}f_{VW}(v,w)&=\frac{1}{2\pi}\exp\left[-\frac{1}{2}[(2v-5w)^2+(-v+3w)^2]\right]\\ &=\frac{1}{2\pi}\exp\left[-\frac{1}{2}(5v^2-26vw+34w^2)\right]\end{aligned}$$

可知，非相关高斯随机变量通过变换成为了相关的高斯随机变量。

例 3.4-3　（两个有序的随机变量）已知两个独立同分布随机变量，概率密度函数为 $f_{X_1}(x) = f_{X_2}(x) = f_X(x)$。定义两个新的随机变量 $Y_1 = \min(X_1, X_2)$ 和 $Y_2 = \max(X_1, X_2)$。很显然，$Y_1 < Y_2$ [①]意味着 Y_1 的实现总比 Y_2 的实现小。根据下面的条件，试求 Y_1，Y_2 的联合概率密度函数 $f_{Y_1Y_2}(y_1, y_2)$。

$$Y_1 = g(X_1, X_2) = \min(X_1, X_2)$$

$$Y_2 = h(X_1, X_2) = \max(X_1, X_2)$$

解　从图 3.4-4 可知（为方便起见，我们只给出了第一象限，但在具体计算中四个象限都要考虑），有两个不相交的实数区域，因此就有两个解。注意到在区域 $\mathscr{R}_1$ 中 $x_1 > x_2$，在区域 $\mathscr{R}_2$ 中 $x_1 < x_2$。因此在 $\mathscr{R}_1$ 中有 $y_1 = x_2$，$y_2 = x_1$，采用 h 和 g 记号就是 $y_1 = g_1(x_1, x_2) = x_2$，$y_2 = h_1(x_1, x_2) = x_1$。该变换的雅可比值为 1，所以 $f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}(y_2, y_1) = f_X(y_2)f_X(y_1)$，$y_1 < y_2$。

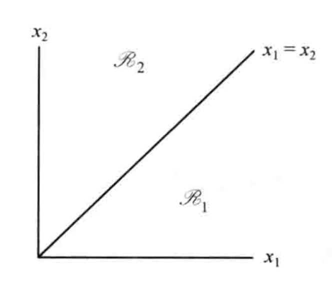


图 3.4-4　例 3.3-6 中讨论的两个区域

在区域 $\mathscr{R}_2$ 中重复上面的计算过程，可得 $y_1 = g_2(x_1, x_2) = x_1$，$y_2 = h_2(x_1, x_2) = x_2$，该变换的雅可比值仍然为 1。因此可得 $f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}(y_1, y_2) = f_X(y_1)f_X(y_2)$，$y_1 < y_2$。把两个区域得到的结果（或根）合并起来

$$f_{Y_1Y_2}(y_1, y_2) = \begin{cases} 2f_X(y_1)f_X(y_2), -\infty < y_1 < y_2 < \infty \\ 0, \text{其他} \end{cases}$$

留给读者的问题：已知 X_1 和 X_2 是相互独立的，那么 Y_1 和 Y_2 也是相互独立的吗？

例 3.4-4　（有序随机变量的边缘概率）在上一个例题中，我们利用两个独立同分布的随机变量 X_1 和 X_2 生成了 Y_1 和 Y_2，$Y_1 < Y_2$。Y_1 和 Y_2 的联合概率密度函数为 $f_{Y_1Y_2}(y_1, y_2) = 2f_X(y_1)f_X(y_2)$，其中 $-\infty < y_1 < y_2 < \infty$。在此试求 Y_1，Y_2 的边缘概率密度函数。

解　欲求 $f_{Y_1}(y_1)$，需要在 $y_2 > y_1$ 区间对 $f_{Y_1Y_2}(y_1, y_2) = 2f_X(y_1)f_X(y_2)$ 进行积分。

$$f_{Y_1}(y_1) = 2f_X(y_1)\int_{y_1}^{\infty} f_X(y_2)\mathrm{d}y_2 = 2f_X(y_1)(1 - F_X(y_1)),\ -\infty < y_1 < \infty$$

类似地，欲求 $f_{Y_2}(y_2)$，需要在区间 $y_1 < y_2$ 对 $f_{Y_1Y_2}(y_1, y_2) = 2f_X(y_1)f_X(y_2)$ 进行积分。

$$f_{Y_2}(y_2) = 2f_X(y_2)\int_{-\infty}^{y_2} f_X(y_1)\mathrm{d}y_1 = 2f_X(y_2)F_X(y_2),\ -\infty < y_2 < \infty$$

例 3.4-5　（两个高斯随机变量的最大值和最小值）我们希望看看用两个高斯随机变量产生新的随机变量，它的概率密度函数是什么样子？为此，令 X_1 和 X_2 为服从正态分布 $N(0, 1)$ 的独立随机变量，定义 $Y_1 \triangleq \min(X_1, X_2)$ 和 $Y_2 \triangleq \max(X_1, X_2)$。根据例 3.4-3[②] 的结论，我们把得到的概率密度函数和正态分布画在同一个坐标系内。图 3.4-5 的曲线是通过 Microsoft Excel[③] 软件实现的。读者可能希望复制这些曲线。

---

① 在此不考虑 $Y_1 = Y_2$ 的情况。

② 原文为例 3.4-5，疑有误——译者注。

③ Excel 是 Microsoft Office 中的软件。关于 Excel 的用法请参考该软件的帮助文档。

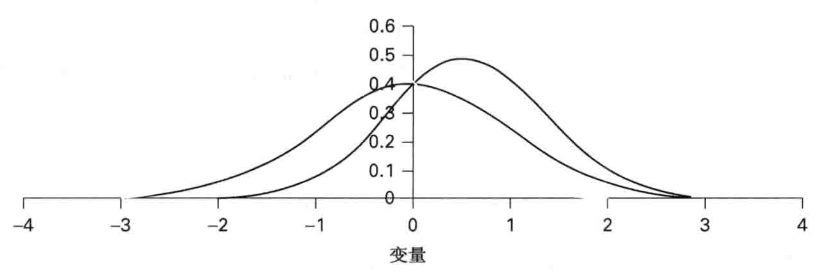


图3.4-5 $\min(X_1, X_2)$ 的概率密度函数的顶点在原点左边的 −0.5 处，$\max(X_1, X_2)$ 的概率密度函数的顶点在原点右边的0.5处。注意到 $\mathrm{Var}[\min(X_1, X_2)] = \mathrm{Var}[\max(X_1, X_2)] < 1$

3.5 补充例子

为了让读者更深入的理解3.4节中介绍的方法，我们在此给出一些补充例题。

例3.5-1 (幅度和角度)考虑如下的随机变量

$$V \triangleq g(X,Y) = \sqrt{X^2+Y^2} \tag{3.5-1}$$

$$W = h(X,Y) = \begin{cases} \arctan\left(\dfrac{Y}{X}\right), & X>0 \\ \arctan\left(\dfrac{Y}{X}\right)+\pi, & X<0 \end{cases} \tag{3.5-2a}$$

随机变量 V 称为幅度或包络，W 称为相位。写成式(3.5-2a)的形式，是因为要在一个 2π 区间上求 w 的解，而且逆函数 $\arctan(y/x)$ 的值域为 $(-\pi/2, \pi/2)$ (即该函数的主值)。

求下面方程的根

$$v = \sqrt{x^2+y^2},\ w = \begin{cases} \arctan\left(\dfrac{y}{x}\right), & x \geqslant 0 \\ \arctan\left(\dfrac{y}{x}\right)+\pi, & x<0 \end{cases} \tag{3.5-2b}$$

我们发现当 $x \geqslant 0$ 时，有 $-\dfrac{\pi}{2} \leqslant w \leqslant \dfrac{\pi}{2}$ 和 $\cos w > 0$。类似地，当 $x<0$ 时，$\dfrac{\pi}{2} < w < \dfrac{3\pi}{2}$ 和 $\cos w < 0$。因此式(3.5-2b)的唯一解是

$$x = v\cos w \triangleq \phi(v,w)$$

$$y = v\sin w \triangleq \psi(v,w)$$

雅可比值 $\tilde{J}$ 为

$$\tilde{J} = \frac{\partial(\phi,\psi)}{(v,w)} = \begin{vmatrix} \cos w & -v\sin w \\ \sin w & v\cos w \end{vmatrix} = v$$

根据式(3.4-11)可知，该方程的解为

$$f_{VW}(v,w) = v f_{XY}(v\cos w, v\sin w) \tag{3.5-3}$$

如果假设 X 和 Y 是独立同分布的，其分布函数为 $N(0, \sigma^2)$，即

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2}\mathrm{e}^{-[(x^2+y^2)/2\sigma^2]}$$

根据式(3.5-3)可知

$$f_{VW}(v,w)=\begin{cases}\left(\dfrac{v}{\sigma^2}e^{-v^2/2\sigma^2}\right)\dfrac{1}{2\pi}, & v>0,-\dfrac{\pi}{2}\leqslant w<\dfrac{3\pi}{2}\\ 0, & \text{其他}\end{cases} \tag{3.5-4}$$

$$=f_V(v)f_W(w)$$

因此，V 和 W 是相互独立的随机变量。包络 V 服从瑞利分布，相位 W 服从 2π 区间上的均匀分布。

例 3.5-2 （幅度和比例）对上一个例题进行适当的改动。设 $V\triangleq\sqrt{X^2+Y^2}$，$W\triangleq Y/X$，根据变换 $g(x,y)=\sqrt{x^2+y^2}$ 和 $h(x,y)=y/x$ 可得下面的方程组

$$v-g(x,y)=0$$
$$w-h(x,y)=0$$

方程组具有两个解

$$x_1=v(1+w^2)^{-1/2},\qquad y_1=wx_1$$
$$x_2=-v(1+w^2)^{-1/2},\qquad y_2=wx_2$$

此时 $-\infty<w<\infty$，$v>0$。当 $v<0$ 时没有实根。

直接可以得出雅可比值为 $|J_1|=|J_2|=(1+w^2)/v$。因此可得

$$f_{VW}(v,w)=\frac{v}{1+w^2}[f_{XY}(x_1,y_1)+f_{XY}(x_2,y_2)]$$

其中 $f_{XY}(x,y)$ 由下式给出

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma^2}\exp[-(x^2+y^2)/2\sigma^2]$$

最终可得

$$f_{VW}(v,w)=\frac{v}{\sigma^2}e^{-v^2/2\sigma^2}u(v)\cdot\frac{1/\pi}{1+w^2}$$
$$=f_V(v)f_W(w)$$

因此，随机变量 V 和 W 是相互独立的，V 服从瑞利分布（同例 3.5-1），W 服从柯西分布。

例 3.5-3 （坐标旋转）设 θ 为一个预定的角度，试考虑下面的旋转变换

$$\begin{aligned}V&\triangleq X\cos\theta+Y\sin\theta\\ W&\triangleq X\sin\theta-Y\cos\theta\end{aligned} \tag{3.5-5}$$

其中 X 和 Y 是独立同分布的高斯随机变量

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma^2}e^{-[(x^2+y^2)/2\sigma^2]}$$

方程组唯一解为

$$v=x\cos\theta+y\sin\theta$$
$$w=x\sin\theta-y\cos\theta$$

也就是

$$x=v\cos\theta+w\sin\theta$$
$$y=v\sin\theta-w\cos\theta$$

雅可比值 $\tilde{J}$ 为

$$\begin{vmatrix} \dfrac{\partial x}{\partial v} & \dfrac{\partial x}{\partial w} \\ \dfrac{\partial y}{\partial v} & \dfrac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{vmatrix} = -1$$

因此有

$$f_{VW}(v,w) = \frac{1}{2\pi\sigma^2}\mathrm{e}^{-[(v^2+w^2)/2\sigma^2]}$$

通过式(3.5-5)的旋转变换 $V = g(X, Y)$ 和 $W = h(X, Y)$ 后，得到的 V 和 W 同 X 和 Y 一样也是独立同分布的高斯随机变量。如果 X 和 Y 是非独立的高斯随机变量，只要 X 和 Y 的联合概率密度函数是高斯(正态)的，那么仍然可能找到一个变换①使得 V 和 W 是独立的高斯随机变量。

例3.5-4 我们重新考虑如何计算例3.3-11中 $Z = \sqrt{X^2 + Y^2}$ 的概率密度函数，这次采用式(3.4-11)来间接计算 $f_Z(z)$。首先需要注意 $Z = \sqrt{X^2 + Y^2}$ 是两个随机变量的一个函数，而式(3.4-11)针对的是两个随机变量的两个函数。为了让二者一致，我们引入一个辅助变量 $W \triangleq X$，使得

$$Z \triangleq g(X,Y) = \sqrt{X^2+Y^2}$$
$$W \triangleq h(X,Y) = X$$

当 $|w| < z$ 时，下面的方程组有两个实根

$$z - g(x,y) = 0$$
$$w - h(x,y) = 0$$

分别为

$$x_1 = w, \qquad x_2 = w$$
$$y_1 = \sqrt{z^2-w^2}, \qquad y_2 = -\sqrt{z^2-w^2}$$

对于这两个根，雅可比值 $|\tilde{J}|$ 都相同

$$|\tilde{J}_1| = |\tilde{J}_2| = \frac{z}{\sqrt{z^2-w^2}}$$

直接套用式(3.4-11)可得

$$f_{ZW}(z,w)\frac{z}{\sqrt{z^2-w^2}}[f_{XY}(x_1,y_1) + f_{XY}(x_2,y_2)]$$

此时假设

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2}\mathrm{e}^{-[(x^2+y^2)/2\sigma^2]}$$

因为 $f_{XY}(x, y) = f_{XY}(x, -y)$，可得

$$f_{ZW}(z,w) = \begin{cases} \dfrac{1}{\pi\sigma^2}\dfrac{z}{\sqrt{z^2-w^2}}\mathrm{e}^{-z^2/2\sigma^2}, & z>0, |w|<z \\ 0, & \text{其他} \end{cases}$$

但是我们并不需要 $f_{ZW}(z, w)$，只需要边缘概率密度函数 $f_Z(z)$。为此，采用式(2.6-47)(用 z

① 参见第5章有关随机向量的内容。

代替 x，w 代替 y)，得到

$$f_Z(z)=\int_{-\infty}^{\infty}f_{ZW}(z,w)\mathrm{d}w$$
$$=\frac{z}{\sigma^2}\mathrm{e}^{-z^2/2\sigma^2}\left[\frac{2}{\pi}\int_0^z\frac{\mathrm{d}w}{\sqrt{z^2-w^2}}\right]u(z)$$

其中，上式方括号里面的值为 1，可用图 3.5-1 来进行验证。令 $w\triangleq z\sin\theta$，那么 $\mathrm{d}w=z\cos\theta\mathrm{d}\theta$，$[z^2-w^2]^{1/2}=z\cos\theta$，此时方括号中的表达式变为

$$\frac{2}{\pi}\int_0^z\frac{\mathrm{d}w}{\sqrt{z^2-w^2}}=\frac{2}{\pi}\int_0^{\pi/2}\mathrm{d}\theta=1$$

因此有

$$f_Z(z)=\frac{z}{\sigma^2}\mathrm{e}^{-z^2/2\sigma^2}u(z)$$

得到的结论与式(3.3-33)相同，不过这里用的是不同的方法。

例 3.5-5　(和与差)最后，我们再次计算例 3.4-1 中的和与差的问题。

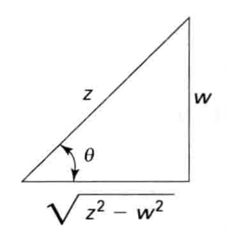


图 3.5-1　三角变换 $w=z\sin\theta$

$$V\triangleq X+Y$$
$$W\triangleq X-Y$$

唯一的解为

$$v-(x+y)=0$$
$$w-(x-y)=0$$

也就是

$$x=\frac{v+w}{2}$$
$$y=\frac{v-w}{2}$$

其中 $|\tilde{J}|=\frac{1}{2}$。所以有

$$f_{VW}(v,w)=\frac{1}{2}f_{XY}\left(\frac{v+w}{2},\frac{v-w}{2}\right)$$

根据概率密度函数的性质

$$f_V(v)=\int_{-\infty}^{\infty}f_{VW}(v,w)\mathrm{d}w$$
$$=\int_{-\infty}^{\infty}\frac{1}{2}f_{XY}\left(\frac{v+w}{2},\frac{v-w}{2}\right)\mathrm{d}w,\qquad z\triangleq\frac{v+w}{2}$$
$$=\int_{-\infty}^{\infty}f_{XY}(z,v-z)\mathrm{d}z$$

上面这个关于两个随机变量之和的重要结论，在 3.3 节[参见式(3.3-12)]中已经通过不同的方法推导出来。

小结

本章主要讨论了随机变量的函数，该部分内容在科学研究，工程应用以及随机过程的学习中有重要的作用。最基本的问题就是如何计算系统输出的概率规律，其中 X 是输入的随机变量，Y 是输出的随机变量，g 是传递函数[也即是 $Y = g(X)$]。此后将问题扩展到两个输入两个输出的情况，其中 X 和 Y 是输入的随机变量，$V = g(X, Y)$ 和 $W = h(X, Y)$ 是输出的随机变量，g 和 h 是传递函数。此时就需要根据 X 和 Y 的联合概率密度函数来计算 V 和 W 的联合概率密度函数。

我们介绍了至少有两种方法可以用来计算大部分关于随机变量的函数：

1. 通过累积分布函数来计算，即所谓的间接法。
2. 通过一种“转动把柄”的直接法。

还介绍了许多关于随机变量转换的问题，包括计算两个随机变量之和的概率密度函数(以及累积分布函数)。这个问题在科学和工程上应用很广泛，比如一个“讨厌的”加性噪声污染了一个希望得到的信号或者测量。例如，所谓的“信号 + 加性噪声”问题就是通信工程中的一个非常根本的问题。

接下来，我们将两个独立随机变量之和的问题扩展到 n 维的情况。随机变量之和的累积分布函数“看起来像”正态分布，从这点入手学习所谓的中心极限定理，这部分内容将在第 4 章讨论。

最后，考虑如何计算两个有序随机变量的累积分布函数。对此我们可以采用一种称为直接法的方法，根据其他随机变量的分布来计算随机变量的概率分布。在第 5 章关于随机向量的内容，我们将要讨论 n 个随机变量的转换问题。有序随机变量体现了统计学的一个分支，即非参量统计，这部分内容经常会得出原始分布独立性的结论。为此，有序随机变量在某种程度上可以得出有关鲁棒性的表达式。

习题

(带 * 号的问题带有一定的难度，需要花更多精力对教材进一步的学习。)

3.1 设已知 X 的累积分布函数为 $F_X(x)$，$Y = aX + b$，$a < 0$。试证明如果 X 不是一个连续随机变量，那么式(3.2-3)应该修改为如下形式：

$$F_Y(y) = 1 - F_X\left(\frac{y-b}{a}\right) + P\left[X = \frac{y-b}{a}\right]$$
$$= 1 - F_X\left(\frac{y-b}{a}\right) + P_X(\frac{y-b}{a})$$

3.2 设 Y 是随机变量 X 的函数，表达式如下：

$$Y \triangleq \begin{cases} X, & X \geqslant 0 \\ X^2, & X \leqslant 0 \end{cases}$$

已知 $f_X(x)$ 如下，求 $f_Y(y)$ 的表达式。在此假设 X: $N(0, 1)$。

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2}$$

3.3 (随机变量的函数)设随机变量 X 服从高斯分布 $N(0,25)$ 。通过函数 g 定义随机变量 $Y = g(x)$

$$g(x) = \begin{cases} 2x & x \geqslant 0 \\ -x & x < 0 \end{cases}$$

请计算 Y 的概率密度函数 $f_Y(y)$。

3.4　设 Y 是随机变量 X 的函数，表达式如下：

$$Y \triangleq \begin{cases} X, & X \geqslant 0 \\ 2X^2, & X < 0 \end{cases}$$

已知 $f_X(x)$ 如下，求 $f_Y(y)$ 的表达式。假设 X: $N(0, 2)$。

$$f_X(x) = \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{4}x^2}$$

3.5　设 X 的概率密度函数为

$$f_X(x) = \alpha e^{-\alpha x} u(x)$$

试计算 Y 的概率密度函数表达式：(a) $Y = X^3$；(b) $Y = 2X + 3$。

3.6　设 X 是一个服从拉普拉斯分布的随机变量，其概率密度函数如下：

$$f_X(x) = \frac{1}{2} e^{-|x|}, \qquad -\infty < x < +\infty$$

令 $Y = g(X)$，其中 $g(\cdot)$ 是一个非线性的饱和限幅器

$$g(x) \triangleq \begin{cases} -1, & x < -1 \\ x, & -1 \leqslant x \leqslant +1 \\ +1, & x > +1 \end{cases}$$

试计算累积分布函数 $F_Y(y)$。

*3.7　在医学成像中，以 X 线断层摄影术为例，探测器读数 y 和人体吸收率 x 之间满足 $y = e^x$ 的规律。令 X: $N(\mu, \sigma^2)$，计算 Y 的概率密度函数。此时 Y 服从对数正态分布。研究发现，正态分布的随机变量在对半导体故障率的数学建模中非常有用。

*3.8　在上一个习题中，你发现如果 X 服从正态分布 X: $N(\mu, \sigma^2)$，那么 $Y \triangleq \exp X$ 服从对数正态分布，其概率密度函数如下：

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left[-\frac{(\ln y - \mu)^2}{2\sigma^2}\right] u(y)$$

(a) 给定一系列的 μ 和 σ 值，试画出对数正态分布的曲线。

(b) 服从对数正态分布的随机变量 Y，试求它的累积分布函数。请将解答写成 erf 函数的形式。提示：有两种可能的求解方法。你可以尝试对上面的密度函数进行积分，或者将这个问题当成随机变量的转换，直接求出 Y 的累积分布函数。

*3.9　在同态成像中，通过对图形函数的非线性变换实现图像增强。假设图像函数是一个随机变量 X，增强后的图像 Y 为 $Y = \ln X$，且 X 不能为负。假设 X 服从指数分布 $f_X(x) = \frac{1}{3} e^{-\frac{1}{3}x} u(x)$，试求 Y 的概率密度函数。

3.10　假设 X: $N(0, 1)$，且设 Y 的定义如下，试计算 Y 的概率密度函数。

$$Y = \begin{cases} \sqrt{X}, & X \geqslant 0 \\ 0, & X < 0 \end{cases}$$

3.11　(a) 假设 X: $N(0, 1)$，令 $Y \triangleq g(X)$，传递函数 g 如图 P3.11 所示，已知 $f_X(x)$，请用间接法计算 $F_Y(y)$ 和 $f_Y(y)$。(b) 请用式(3.2-23)计算 $f_Y(y)$。请问在 $y = 0, 1$ 时，为什么不能用式(3.2-23)来计算 $f_Y(y)$？

3.12　设 X: $U[0, 2]$，计算 $Y = g(X)$ 的概率密度函数，其中函数 g 如图 P3.12 所示。

3.13　设 X: $U[0, 2]$，计算 $Y = g(X)$ 的概率密度函数，其中函数 g 如图 P3.13 所示。

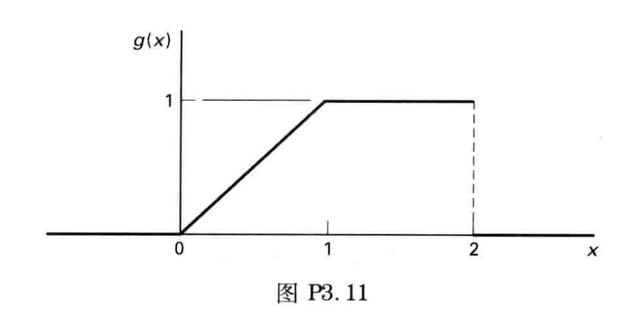


图 P3.11

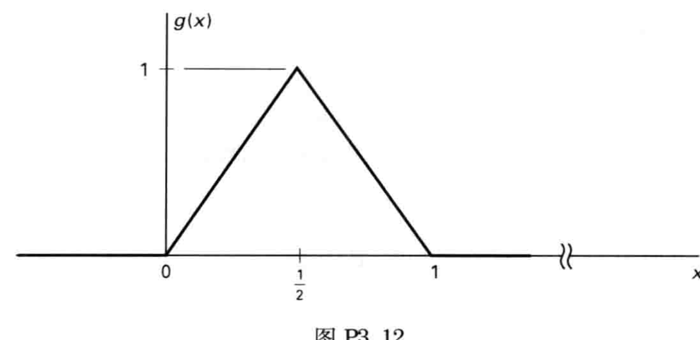


图 P3.12

3.14 设随机变量 X 服从高斯分布 X: $N(0, 1)$，其概率密度函数如下：

$$f_X(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{-x^2}{2}},\qquad -\infty<x<+\infty$$

令 $Y = g(X)$，其中 g 是一个非线性函数，称为饱和限幅器，其表达式如下：

$$g(x)\triangleq\begin{cases}-1, & x<-1\\ x, & -1\leqslant x\leqslant 1\\ 1, & x>1\end{cases}$$

(a)绘出 $g(x)$ 的图像；(b)计算 $F_Y(y)$；(c)计算 $f_Y(y)$ 并画出它的图像。

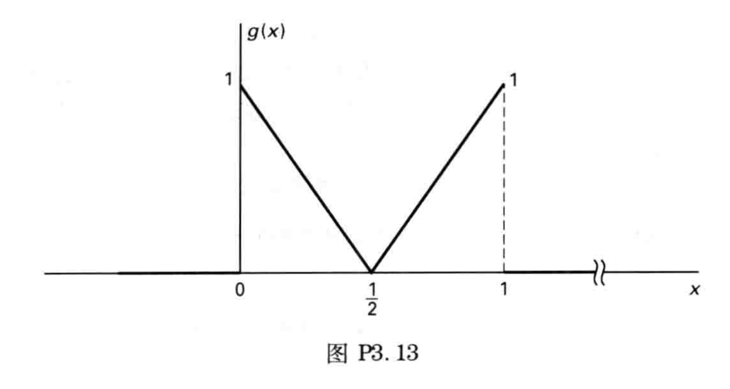


图 P3.13

3.15 请计算 $Y = a/X(a > 0)$ 的概率密度函数。请证明：如果 X 服从参数为 α 的柯西分布，那么 Y 服从参数为 α/a 的柯西分布。

3.16 设 $Y \triangleq \sec X$，试用 $f_X(x)$ 表示 $f_Y(y)$。如果 $f_X(x)$ 是区间 $(-\pi, \pi]$ 上的均匀分布，试求此时的 $f_Y(y)$。

3.17 对于两个服从均匀分布的随机变量 X:$(-1, +1)$ 和 Y:$(-2, +2)$，当 X 和 Y 相互独立时，请计算 $Z = X + Y$ 的概率密度函数。

3.18 设 X 和 Y 是独立同分布的随机变量，二者都服从指数分布

$$f_X(x)=f_Y(x)=\alpha \mathrm{e}^{-\alpha x}u(x)$$

请计算 $Z \triangleq X-Y$ 的概率密度函数。

3.19 已知随机变量 X 和 Y 的联合概率密度函数为 $f_{X,Y}(x,y)$。定义新的随机变量如下：

$$V \triangleq X+Y \quad \text{和} \quad W \triangleq 2X-Y$$

(a)试用联合概率密度函数的表达式 $f_{X,Y}(x,y)$ 表示 $f_{V,W}(v,w)$。

(b)请证明：根据(a)的结论或者其他合理的结论，对于 $Z \triangleq X+Y$，要在什么条件下等式才可以成立？

$$f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\mathrm{d}x$$

3.20 当 $f_X(x)=\mathrm{e}^{-x}u(x)$ 时，请再次求解例3.2-11。

3.21 当 $f_X(x)=\mathrm{e}^{-x}u(x)$ 时，请再次求解例3.2-12。

3.22 如果希望产生一组随机数，其概率密度函数如图P3.22所示。我们唯一可用的就是(0, 1)区间上的均匀随机数发生器，请问该如何处理？

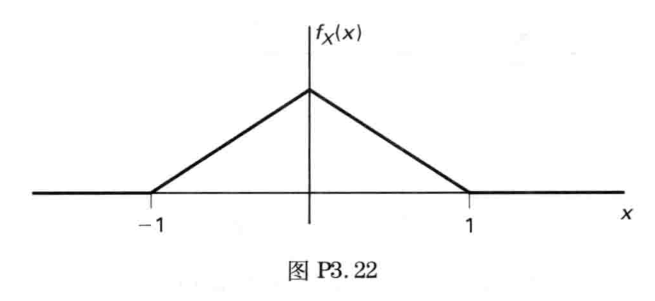


图 P3.22

3.23 如果希望产生一组零均值的高斯随机数。唯一可用的仍是(0, 1)区间上的均匀随机数发生器，请问该如何处理？建议把12个均匀分布的随机数相加，再把得到的结果减去6，就可以得到高斯分布随机数。请编写一个程序产生10 000个随机数，并画出结果的直方图。直方图是一种柱状图，x 轴上有一系列的矩形，y 轴表示落在这个矩形范围内的点数。请从 −10 ~ 10 的区间上画200个矩形，每个矩形的宽度为0.1。请问在直方图的哪个区间看起来“最像”高斯分布，哪个区间看起来“最不像”？同时请解释一下为什么用这个方法可以产生高斯分布随机数？

*3.24 一般而言，计算机只能产生均匀分布的随机数 X: $U[0,1]$。现在的问题是：如何对 X 进行适当的变换产生任意分布的随机数？

(a)拉普拉斯分布经常在图形处理问题中遇到，当概率密度函数为 $f_Y(y)=\frac{c}{2}\exp(-c|y|)$，$-\infty<y<+\infty$，$c>0$。请求出拉普拉斯分布的累积分布函数 $F_Y(y)$，$-\infty<y<+\infty$。

(b)考虑如下的变换：

$$z=g(x)=F_Y^{-1}(x)$$

请用(a)中得到的累积分布函数表达式，其中 F_Y^{-1} 表示逆函数。证明当 X: $U[0,1]$ 时，$Z=g(X)$ 服从参数为 c 的拉普拉斯分布。请注意：只要逆函数存在，结果并不依赖于拉普拉斯分布函数。

(c)这种变换方法有没有什么限制条件？可以适用于混合随机变量吗？如果分布函数有平坦区域可以吗？适用于离散随机变量吗？

3.25 计算习题3.18中 $|Z|$ 的概率密度函数。

3.26 设 X 和 Y 是独立的连续随机变量，令 $Z=\min(X,Y)$，试计算 $F_Z(z)$ 和 $f_Z(z)$。如果 X 和 Y 的分布为 $U(0,1)$，请绘出 $F_Z(z)$ 和 $f_Z(z)$ 的图像。当概率密度函数为 $f_X(x)=f_Y(x)=\alpha\exp[-\alpha x]\cdot u(x)$ 时，请重复上面的计算。

3.27 设 $Z \triangleq \max(X_1,X_2)$，其中 X_1 和 X_2 是独立的指数分布随机变量，概率密度函数为

$$f_{X_i}(x)=\frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right)u(x), \quad i=1,2$$

(a)请计算 Z 的分布函数。

(b)请计算 Z 的概率密度函数。

3.28 对于 n 个独立同分布(i. i. d.)的随机变量 $X_1,X_2,\cdots,X_n$，累积分布函数都为 $F_{X_i}(x) \triangleq F(x)$ 。设 $Z \triangleq \max[X_1,X_2,\cdots,X_n]$，请用 $F(x)$ 来表示 Z 的累积分布函数。

3.29 对于 n 个独立同分布(i. i. d.)的随机变量 $X_1,X_2,\cdots,X_n$，累积分布函数都为 $F_{X_i}(x) \triangleq F(x)$ 。设 $Z \triangleq \min[X_1,X_2,\cdots,X_n]$，请用 $F(x)$ 来表示 Z 的累积分布函数。

3.30 设 X_1, X_2, $\cdots$, X_n 是 n 个独立同分布的随机变量，概率密度函数为 $f_{X_i}(x) = e^{-x}u(x)$。计算 $Z_n = \max(X_1, X_2, \cdots, X_n)$ 的概率密度函数，并画出 $n=3$ 时的图像。

设 X_1, X_2, $\cdots$, X_n 是 n 个独立同分布的随机变量，概率密度函数为 $f_{X_i}(x) = e^{-x}u(x)$。计算 $Z_n = \min(X_1, X_2, \cdots, X_n)$ 的概率密度函数，并画出 $n=3$ 时的图像。

3.31 设 X 和 Y 是独立同分布的随机变量，服从 $U(-1, 1)$。计算并画出图 P3.31 中 Z 的概率密度函数。在这里只能对正数进行求平方根操作，否则输出是 0。

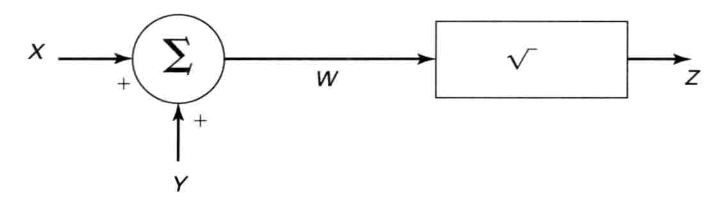


图 P3.31 求平方根设备

3.32 一架飞机可以飞行的时间 Z 用 $Z = \alpha X$ 来建模，其中 X 是油箱中的油量，α 是一个正的比例常数。假设在一架飞机上有两个独立的油箱，使得一个油箱耗尽的时候可以自动启用另外一个油箱中的燃料。由于管理松懈的原因，使得飞机起飞时两个油箱的油量都没有检查过。用 X_1 表示第一个油箱中的油量，X_2 表示第二个油箱中的油量，且 X_1 和 X_2 是独立同分布的随机变量，概率密度函数为 $f_{X_1}(x) = f_{X_2}(x) = \frac{1}{b}[u(x) - u(x-b)]$。用 Z 表示飞机能够飞行的最长时间，请计算 Z 的概率密度函数。如果 $b=100$(单位：升)，$\alpha = 1$（单位：小时/10 升)，请问飞机至少能够飞行 5 个小时的概率为多少？

3.33 设 X 和 Y 为两个独立同分布的泊松随机变量，概率质量函数为

$$P_X(k) = \frac{1}{k!}e^{-2}2^k u(k)$$

$$P_Y(k) = \frac{1}{k!}e^{-3}3^k u(k)$$

请计算 $P[Z \leqslant 5]$ 的概率，其中 $Z \triangleq X + Y$（提示：$\sum_{j=0}^{n}\binom{n}{j}a^j b^{n-j} = (a+b)^n$）。

3.34 已知随机变量 X 和 Y 是独立同分布的均匀随机变量，服从 $U(0, 1)$ ：

(a)定义随机变量 U 和 V 如下，请计算它们的联合概率密度函数 $f_{U,V}$

$$U \triangleq \frac{1}{2}(X+Y)$$

$$V \triangleq \frac{1}{2}(X-Y)$$

(b)请画出 $f_{U,V}$ 在 (u, v) 平面上的支集。函数的支集是指该函数值域上非 0 的子集。

3.35 设 X 和 Y 是独立的均匀分布随机变量，概率密度函数分别为 $f_X(x) = \frac{1}{2}$, $|x| < 1$ 和 $f_X(y) = \frac{1}{4}$, $|y| < 1$ 。请计算：(a) $Z \triangleq X + Y$ 的概率密度函数；(b) $Z \triangleq 2X - Y$ 的概率密度函数。

3.36 计算联合概率密度函数 $f_{ZW}(z, w)$，Z 和 W 的定义如下：

$$Z \triangleq X^2 + Y^2$$
$$W \triangleq X$$

其中

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2} e^{-[(x^2+y^2)/2\sigma^2]}, -\infty < x < \infty, -\infty < y < \infty$$

再根据得到的结果计算 $f_Z(z)$。

*3.37　考虑如下的变换：

$$Z = aX + bY$$
$$W = cX + dY$$

设

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-Q(x,y)}$$

其中

$$Q(x,y) = \frac{1}{2\sigma^2(1-\rho^2)}[x^2 - 2\rho xy + y^2]$$

请问当系数 a, b, c, d 应该取什么值时，Z 和 W 才是独立的高斯随机变量？

3.38　已知 X 和 Y 的联合概率密度函数

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\left(\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right)\right]$$

计算 V 和 W 的联合概率密度函数 $f_{VW}(v, w)$，V 和 W 的定义如下：

$$V = \tfrac{1}{2}(X^2 + Y^2)$$
$$W = \tfrac{1}{2}(X^2 - Y^2)$$

3.39　请根据式(3.4-11)或式(3.4-12)，根据直接法来推导式(3.4-4)。

3.40　考虑如下的变换：

$$Z = X\cos\theta + Y\sin\theta$$
$$W = X\sin\theta - Y\cos\theta$$

请计算 $f_{ZW}(z, w)$，已知 $f_{XY}(x, y)$ 表达式如下：

$$f_{XY}(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}, -\infty < x < \infty, -\infty < y < \infty$$

(可以把这个变换看成对 Z 进行了 $+\theta$ 的旋转，而 W 是反方向的旋转)

3.41　计算如下表达式的联合概率密度函数：

$$Z \triangleq X^2 + Y^2$$
$$W \triangleq 2Y$$

已知 $f_{XY}(x, y)$ 如下：

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2} e^{-[(x^2+y^2)/2\sigma^2]}$$

3.42　设 $f_{XY}(x,y) = A(x^2 + y^2)$，$0 \leqslant x \leqslant 1$，$|y| \leqslant 1$，其余情况 $f_{XY}(x,y)$ 为 0。请对所有的 x 和 y 计算累积分布函数 $F_{XY}(x,y)$，并给出 A 的合理取值。

3.43　考虑 3.1 节中的输入-输出问题，设随机变量 X 是系统的输入，系统输出为 $Y = g(X)$。

(a) Y 的定义域是什么？

(b) X 和 Y 合理的概率空间是什么？

(c) 请问 R^1 的哪个子集包含事件 $\{Y \leqslant y\}$？

(d) 如果 $Y = 2X + 3$，对于 Y 而言，事件 $(-\infty, y]$ 的逆像是什么？

3.44 如图 P3.44 所示的框图，该系统把信号 X 从点 a 传递到点 b。传递链路 L_1 和 L_2 独立运行，二者的故障时间分别为 T_1 和 T_2。T_1 和 T_2 独立同分布，且服从参数为 $\lambda(>0)$ 的指数分布。用 Y 表示输出，如果 $Y=0$ 表示两条链路同时出现故障。请计算 Y 在 t 时刻的累积分布函数 $F_Y(y, t)$，证明无论 t 取什么值，都有 $F_Y(\infty, t)=1$。

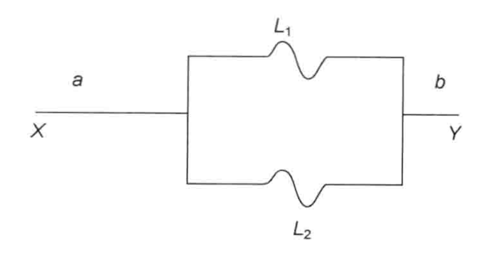


图 P3.44 并行级联

参考文献

3-1 W. F. Davenport, *Probability and Random Processes: An Introduction for Applied Scientists and Engineers*. New York: McGraw-Hill, 1970.

3-2 H. Stark, F. B. Tuteur, and J. B. Anderson, *Modern Electrical Communications*. 2nd edition, Upper Saddle River, N. J.: Prentice-Hall, 1988.

3-3 S. O. Rice, "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, Vols. 23, 24, 1944, 1945.

3-4 J. Marsden and A. Weinstein, *Calculus*. Menlo Park, CA.: Benjamin/Cummings, 1980.

3-5 A. Papoulis, and S. U. Pillai *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 4th Ed, 2002.

补充文献

Cooper, G. R. and C. D. McGillem, *Probabilistic Methods of Signal and System Analysis*, 3rd edition. New York: Holt, Rinehart and Winston, 1999.

Leon-Garcia, A., *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd edition. Reading, MA: Prentice Hall, 2008.

Helstrom, C. W., *Probability and Stochastic Processes for Engineers*, 2nd edition. New York: Macmillan, 1991.

Papoulis, A., *Probability &Statistics*. Englewood Cliffs, NJ: Prentice Hall, 1990.

Peebles, P. Z. Jr., *Probability, Random Variables, and Random Signal Principles*, 4th edition. New York: McGraw-Hill, 2001.

Scheaffer, R. L., *Introduction to Probability and Its Applications*. Belmont, CA: Duxbury, 1990.

Strook, D. W., Probability Theory, an Analytic View, 2nd edition, Cambridge University Press, Cambridge, England 2010.

Viniotis, Y., *Probability and Random Processes for Electrical Engineers*. New York: McGraw-Hill, 1998.

Yates, R. D. and D. J. Goodman, *Probability and Stochastic Processes*, 2nd edition, New York: Wiley, 2004.

Ziemer, R. E., *Elements of Engineering Probability &Statistics*. Upper Saddle River, NJ: Prentice Hall, 1997.

第 4 章　数学期望和矩

4.1　随机变量的数学期望

在许多情况下，我们总是希望用一些数字来归纳一个随机变量的性质或其分布规律。这些数字一般都是各种各样的平均值，或者称为随机变量的数学期望。我们还经常用“矩”(moments)这个名词来描述一大类的平均值，这方面的内容将在后面讲解。

对于一组数字的平均值我们都很熟悉，比如某次考试的平均分，5 岁儿童的平均身高和体重，男性和女性平均寿命等。一般而言，我们采用下面的公式来计算一组数字 $x_1, x_2, \cdots, x_N$ 的平均值

$$\mu_s = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4.1-1}$$

其中下标 s 表示 μ_s 是一个集合(set)的平均值。

一组数字 $x_1, x_2, \cdots, x_N$ 的平均值 μ_s 也可以看成这个集合的“重心”。更准确地说，平均值就是与集合中所有数字最接近的那个值，也就是它与集合中所有数字的距离之和是最小的。为此我们需要求出这个数值 z，使得 z 与所有点的距离之和 D 或者距离之和的平方 D^2 最小。也就是

$$D^2 \triangleq \sum_{i=1}^{N}(z-x_i)^2$$

很显然，只有当 $\mathrm{d}D^2/\mathrm{d}z = 0$ 时才会取到最小值，或者有

$$\frac{\mathrm{d}D^2}{\mathrm{d}z} = 2Nz - 2\sum_{i=1}^{N} x_i = 0$$

可以得出

$$z = \mu_s = \frac{1}{N}\sum_{i=1}^{N} x_i$$

我们注意到，在式(4.1-1)中所有的数字权重都相同(也就是说，每个 x_i 都被乘以了一个相同的因子 $1/N$)。如果出于某种原因，在计算平均值的时候我们希望有些数字的权重大一些，那么此时得到的就是加权平均值。但是，在本章的后续内容中我们不再考虑加权平均这个概念。

根据式(4.1-1)，可以得到一个集合“最具代表性”的数值或“重心”，但这并不能告诉我们集合中的数字与平均值偏离或发散的程度。比如，集合 $S_1 = \{0.9, 0.98, 0.95, 1.1, 1.02, 1.05\}$ 和 $S_2 = \{0.2, -3, 1.8, 2, 4, 1\}$ 平均值是相同的，但很显然 S_2 中的数字比 S_1 中的数字更偏离平均值。用来统计这种偏离程度的平均值就称为集合的标准偏差，用 σ_s 来表示，计算公式如下：

$$\sigma_s = \left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_s)^2\right]^{1/2} \tag{4.1-2}$$

式(4.1-1)和式(4.1-2)虽然很重要，但它们并没有很好地揭示平均值的巨大价值。为了充分

利用平均值的有关特性，有必要推导一种根据概率论来计算平均值的方法。

考虑一个概率空间 $(\Omega, \mathscr{F}, P)$，以及相关的实验 $\mathscr{H}$ 和离散随机变量 X。对于 $\mathscr{H}$ 的每个输出 ζ_i 对应一个数值 $X(\zeta_i) \triangleq x_i$，也就是随机变量 X 的可能取值。设 $x_1, x_2, \cdots, x_M$ 表示 X 的 M 个不同的取值。在此假设实验 $\mathscr{H}$ 被重复了 N 次，设 $x^{(k)}$ 是第 k 次实验观测到的输出，而且 $x^{(k)}$ 一定是 $x_1, x_2, \cdots, x_M$ 中的一个数值。假设在这 N 次实验中，x_1 发生了 n_1 次，x_2 发生了 n_2 次，以此类推。当 N 越来越大时，我们可以就用下面的公式来计算 X 的平均值 μ_X

$$\begin{aligned}\mu_X &\approx \frac{1}{N}\sum_{k=1}^{N} x^{(k)} \\ &\approx \frac{1}{N}\sum_{i=1}^{M} n_i x_i \end{aligned} \tag{4.1-3}$$

$$\approx \sum_{i=1}^{M} x_i \left(\frac{n_i}{N}\right) \tag{4.1-4}$$

$$\approx \sum_{i=1}^{M} x_i P[X = x_i] \tag{4.1-5}$$

例 4.1-1 (灌铅的骰子)我们将一个灌铅的骰子投掷了 17 次。其中 $N=17$，$M=6$(表示这是 6 个面的骰子)。观测到的结果是 $\{1, 3, 3, 1, 2, 1, 3, 2, 1, 1, 2, 4, 1, 1, 5, 3, 6\}$。令 $P[i]$ 表示观测到数字 i 的概率，那么通过上面的观测结果可得

$$P[1] \approx 7/17; P[2] \approx 3/17; P[3] \approx 4/17; P[4] \approx 1/17; P[5] \approx 1/17; P[6] = 1/17$$

通过这种方法来估计“真实的”概率非常不可靠。为了得到可靠的数据，不得不拼命增加投掷的次数。我们可能会问：到底“真实的”概率是什么？一种想法认为必须通过实验来确定概率，因为概率就好像一个自然规律一样根植在这个骰子中。另外一种想法认为当 N 为任意大的取值时，真实的概率就是比值 $P[i] = n_i/N$。何谓“任意大”？对于任意一个有限值 N，当 N 变大时得到的概率估计值总是变化的。这个棘手的问题将会用统计的方法来很好解决，这方面的详细内容详见第 6 章和第 7 章。

式(4.1-5)从频率的角度给出了概率的定义，下面将给出第一个定义。

定义 4.1-1 随机变量 X 的概率质量函数为 $P_X(x_i)$，x_i 是所有可能的取值，则 X 的数学期望(平均值)的定义如下：

$$E[X] \triangleq \sum_i x_i P_X(x_i) \tag{4.1-6}$$

从上式可见，数学期望是在随机变量产生的概率空间中计算出来的。我们也可以把离散样本空间中的所有点都累加起来计算数学期望，即 $E[X] = \sum_{\Omega} X(\zeta_i) P[\{\zeta_i\}]$，其中 ζ_i 是样本空间 Ω 中的离散输出点。

下面将要给出的定义，对于连续和离散的随机变量都适用。

定义 4.1-2 概率密度函数为 $f_X(x)$ 的实随机变量 X，如果它的数学期望(或平均值)存在①，那么定义如下：

---

① 数学期望存在是指在积分区间上绝对可积，即 $\int_{-\infty}^{\infty} |x| f_X(x)\mathrm{d}x < \infty$。

$$E[X]=\int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x \tag{4.1-7}$$

同定义 4.1-1 类似，数学期望可在原始概率空间中计算得到。如果样本描述空间不是离散的而是连续的，例如一条实数直线，这是一个不可数的无限集合，此时 $E[X]=\int_{\Omega} X(\zeta)P[\{\mathrm{d}\zeta\}]$，其中 $P[\{\mathrm{d}\zeta\}]$ 表示无限小事件 $\{\zeta<\zeta'\leqslant\zeta+\mathrm{d}\zeta\}$ 发生的概率。

常用符号 $E[X]$，$\bar{X}$，μ_X 或 μ 来表示 X 的数学期望。对于一个随机变量的方程 $Y=g(X)$，根据式(4.1-7)，Y 的数学期望为

$$E[Y]=\int_{-\infty}^{\infty} y f_Y(y)\mathrm{d}y \tag{4.1-8}$$

可是，式(4.1-8)需要事先用 $f_X(x)$ 计算出 $f_Y(y)$。如果我们只需计算数学期望 $E[Y]$，有没有可能不用先计算出 $f_Y(y)$？定理 4.1-1 将回答这个问题。

定理 4.1-1　$Y=g(X)$ 的数学期望可由如下公式计算得到：

$$E[Y]=\int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x \tag{4.1-9}$$

其中 g 是一个可测(Borel)函数①。式(4.1-9)是概率论中非常重要的一个结论。对式(4.1-9)的严格证明需要用到勒贝格积分的知识，在此我们用一个非正式的自变量来证明式(4.1-9)②。

证明式(4.1-9)③

根据 3.2 节中的结论，如果 $Y=g(X)$，那么对于任意的 y_j 都有(参见图 4.1-1)

$$\{y_j<Y\leqslant y_j+\Delta y_j\}=\bigcup_{k=1}^{r_j}\{x_j^{(k)}<X\leqslant x_j^{(k)}+\Delta x_j^{(k)}\} \tag{4.1-10}$$

其中 r_j 是方程 $y_j-g(x)=0$ 的实根个数，也就是

$$y_j=g(x_j^{(1)})=\cdots=g(x_j^{(r_j)}) \tag{4.1-11}$$

式(4.1-10)中的等号表示对于映射 X 和 Y，基本事件都是相同的。因此，等号两边事件的概率都是相同的。式(4.1-10)中等号右边的事件是不相交的，而且并集的概率就是所有独立事件概率之和。把 y 轴划分为许多子区间 y_1，y_2，…，y_j，…，根据式(2.4-6)以及黎曼(Riemann)④求和，可将式(4.1-8)近似写为⑤

$$\begin{aligned}E[Y]&=\int_{-\infty}^{\infty} y f_Y(y)\mathrm{d}y\\&\approx\sum_{j=1}^{m} y_j P[y_j<Y\leqslant y_j+\Delta y_j]\\&=\sum_{j=1}^{m}\sum_{k=1}^{r_j} g(x_j^{(k)})P[x_j^{(k)}<X\leqslant x_j^{(k)}+\Delta x_j^{(k)}]\end{aligned} \tag{4.1-12}$$

---

① 可测方程的定义参见本书 3.1 节。

② 参见 Feller[4-1, p.5]，或者 Davenport[4-2, p.223]。

③ 原文此处是式(4.1-8)，明显有误——译者注。

④ Bernhard Riemann(1826-1866)，德国数学家，在积分理论方面做出了巨大贡献。

⑤ 自变量的定义参见 Papoulis[4-3, p.141]。

其中，式(4.1-12)最后一行的推导利用了式(4.1-10)和式(4.1-11)。由于 $x_j^{(k)}$ 都是不同的，因此完全可以只用 i 来区分，而没必要用 j 和 k 两个标记索引。此时，式(4.1-12)可以写为

$$E[Y] \approx \sum_{i=1}^{n} g(x_i) P[x_i < X \leqslant x_i + \Delta x_i]$$

当 Δy, $\Delta x \to 0$ 时，可得到精确的解

$$E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x \tag{4.1-13}$$

式(4.1-13)用到了黎曼近似求和以及式(2.4-6)，x_i 按照从小到大的顺序，$x_1 < x_2 < x_3 < \cdots$。

对于 X 是离散随机变量的这种特例

$$E[Y] = \sum_{i} g(x_i) P_X(x_i) \tag{4.1-14}$$

根据式(4.1-13)就可以得到这个结果，这是因为离散随机变量的概率密度函数由一系列的 δ 函数组成，具体性质可以参考附录 B 中的式(B.3-1)。

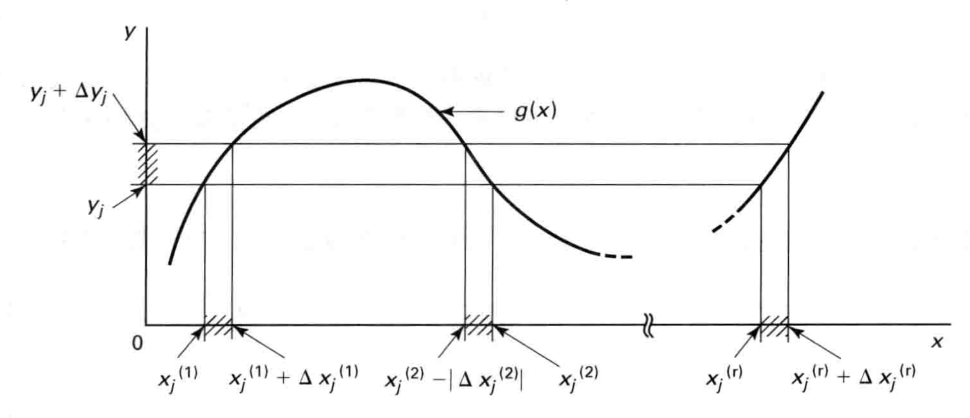


图 4.1-1 式(4.1-10)中事件的等价性

例 4.1-2 (高斯分布的数学期望)设 X: $N(\mu, \sigma^2)$，读为"X 服从参数为 μ 和 σ^2 的正态分布"。X 的数学期望(均值)如下：

$$E[X] = \int_{-\infty}^{\infty} x \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right) \right) \mathrm{d}x$$

设 $z \triangleq (x-\mu)/\sigma$，可得

$$E[X] = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z\mathrm{e}^{-\frac{1}{2}z^2}\mathrm{d}z + \mu\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}z^2}\mathrm{d}z \right)$$

由于奇对称，第一项积分结果为 0；第二项积分结果为 μ，这是因为当 Z: $N(0, 1)$ 时，括号中的 $P[Z \leqslant \infty]$ 是一个必然事件。因此有

$$E[X] = \mu, \qquad X : N(\mu, \sigma^2)$$

因此，$N(\mu, \sigma^2)$ 中的参数 μ 其实就是 X 的数学期望或均值，与 2.4 节得到的结果是一致的。

例 4.1-3 (伯努利随机变量的数学期望)设随机变量 B 服从伯努利分布，取值为 1 的概率为 p，取值为 0 的概率为 $q = 1 - p$。概率质量函数如下：

$$P_B(k) = \begin{cases} p, & k=1 \\ q, & k=0 \\ 0, & \text{其他} \end{cases}$$

那么数学期望由如下计算可得：

$$\begin{aligned} E[B] &= \sum_{b=-\infty}^{+\infty} kP_B(k) \\ &= 1p + 0 \\ &= p \end{aligned}$$

例 4.1-4　（二项分布随机变量的数学期望）设随机变量 K 服从二项分布，概率质量函数为 $P_K(k) = b(k;n,p)$，数学期望可由下式计算：

$$\begin{aligned} E[K] &= \sum_{k=-\infty}^{+\infty} kP_K(k) \\ &= \sum_{k=0}^{n} kb(k;n,p) \\ &= \sum_{k=0}^{n} k\binom{n}{k} p^k(1-p)^{n-k} \\ &= \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!} p^k(1-p)^{n-k} \\ &= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k(1-p)^{n-k} \\ &= \sum_{k'=0}^{n-1} \frac{n!}{k'!(n-k'-1)!} p^{k'+1}(1-p)^{n-k'-1}, \quad k' \triangleq k-1 \\ &= np\left(\sum_{k'=0}^{n-1} \frac{(n-1)!}{k'!(n-1-k')!} p^{k'}(1-p)^{n-1-k'}\right) \\ &= np, \quad \text{括号中的求和值为} \sum_{k=0}^{n-1} b(k;n-1,p) = 1 \end{aligned}$$

例 4.1-5　（更多关于复式彩票的问题）在此我们继续讨论第 1 章的例 1.9-6，看看到底是一次买 50 张彩票好，还是每次只买一张彩票，连续买 50 次好，其中每张彩票的中奖概率是一样的且相互独立。现在我们关心的是每次投注中平均或期望的收益是多少？在每次投注中，有 100 张面值 1 美元的彩票，对于获胜者而言公平的回报是 100 美元。对于单次投注，获奖的概率是 50%，所以预期收益就是 50 美元。对于投注 50 次的情况，获胜的次数 K 服从二项分布 $b(k;50,0.01)$，所以 K 的平均值为 $E[K] = np = 50 \times 0.01 = 0.5$。因为总的收益为 $100K$ 美元。所以平均收益为 50 美元，这与一次买 50 张彩票的情况一样。

例 4.1-6　（泊松分布的数学期望）设 K 是一个服从泊松分布的随机变量，参数 $a > 0$，可得

$$
\begin{aligned}
E[K] &= \sum_{k=0}^{\infty} k \frac{\mathrm{e}^{-a}}{k!} a^k \\
&= a \sum_{k=0}^{\infty} \frac{\mathrm{e}^{-a}}{(k-1)!} a^{k-1} \\
&= a \sum_{k=1}^{\infty} \frac{\mathrm{e}^{-a}}{(k-1)!} a^{k-1} \\
&= a \sum_{i=0}^{\infty} \frac{\mathrm{e}^{-a}}{i!} a^{i} \\
&= a
\end{aligned} \tag{4.1-15}
$$

因此，泊松随机变量的数学期望就是其参数 a。

数学期望的线性特性 如果把取数学期望 E 看成一个运算符，那么很显然这是一个线性运算符。对于任意一个 X 有

$$E\left[\sum_{i=1}^{N} g_i(X)\right] = \int_{-\infty}^{+\infty} \left(\sum_{i=1}^{N} g_i(x)\right) f_X(x)\mathrm{d}x \tag{4.1-16}$$

$$= \sum_{i=1}^{N} \int_{-\infty}^{+\infty} g_i(x) f_X(x)\mathrm{d}x \tag{4.1-17}$$

$$= \sum_{i=1}^{N} E[g_i(X)] \tag{4.1-18}$$

从上式结论可看出这种线性特性。此外对于两个随机变量之和，数学期望运算符 E 也是线性的

$$
\begin{aligned}
E[X+Y] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x+y) f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} x \left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\mathrm{d}y\right) \mathrm{d}x + \int_{-\infty}^{+\infty} y \left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\mathrm{d}x\right) \mathrm{d}y \\
&= \int_{-\infty}^{+\infty} x f_X(x)\mathrm{d}x + \int_{-\infty}^{+\infty} y f_Y(x)\mathrm{d}y \\
&= E[X] + E[Y]
\end{aligned}
$$

读者将会注意到，这个结论还可以拓展到 N 个随机变量 X_1, X_2, $\cdots$, X_N 之和的情况，也就是

$$E\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} E[X_i] \tag{4.1-19}$$

需要注意的是，这个结论并不要求随机变量是相互独立的。对于上面提到的两个线性性质，我们可以归纳为：数学期望运算符 E 满足随机变量的加法分配律。

例 4.1-7 （高斯分布的方差）设 X: $N(\mu, \sigma^2)$ 是均值为 μ 的随机变量，利用数学期望运算的线性特性，方差为 $E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - \mu^2 = \mathrm{Var}[X]$

$$E[(X-\mu)^2]=\int_{-\infty}^{+\infty}(x-\mu)^2\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}z^2\mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z\quad \text{其中需要代入 } z\triangleq(x-\mu)/\sigma$$

接下来，利用分部积分法，设 $u = z$，$\mathrm{d}v = z\mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z$，可得 $\mathrm{d}u = \mathrm{d}z$，$v = e^{-\frac{z^2}{2}}$，因此上述积分表达式可写为

$$\int_{-\infty}^{+\infty}z^2\mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z=\left(-z\mathrm{e}^{-\frac{z^2}{2}}\right)\Big|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z$$

$$=-0+0+\sqrt{2\pi}$$

利用标准正态分布 $N(0, 1)$ 在整个区间的积分为 1，可求出最后一项的积分结果 $E[(X-\mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}}\sqrt{2\pi} = \sigma^2$。因此对于均值为 μ 的高斯随机变量 X，分布函数中的参数 σ^2 就是 X 的方差。

至此，根据高斯密度函数的定义，第 2 章中介绍的那些参数其实是高斯分布的数学期望和方差。在实际操作中，一般是通过对 X 进行大量的独立观测来估计这两个基本参量，对均值的估计要用到式(4.1-1)，对 σ 的估计要用到式(4.1-2)。

例 4.1-8　(柯西分布的数学期望) $\alpha(-\infty<\alpha<\infty)$ 和 $\beta(\beta>0)$ 是柯西分布概率密度函数的两个参量，定义如下：

$$f_X(x)=\frac{1}{\pi\beta\left(1+\left(\frac{x-\alpha}{\beta}\right)^2\right)},\qquad -\infty<x<\infty \tag{4.1-20}$$

设 X 服从柯西分布，且 $\beta = 1$，$\alpha = 0$。那么有

$$E[X]=\int_{-\infty}^{\infty}x\left(\frac{1}{\pi(x^2+1)}\right)\mathrm{d}x$$

上式是一个广义积分，在一般情况下是不收敛的。但是，如果在柯西主值意义下计算积分，可得

$$E[X]=\lim_{x_0\to\infty}\left[\int_{-x_0}^{x_0}x\left(\frac{1}{\pi(x^2+1)}\right)\mathrm{d}x\right] \tag{4.1-21}$$

可得 $E[X] = 0$。此外，如果定义 $Y \triangleq X^2$，那么 $E[Y]$ 无论如何都是不存在的，因为

$$E[Y]=\int_{-\infty}^{\infty}x^2\left[\frac{1}{\pi(x^2+1)}\right]\mathrm{d}x=\infty \tag{4.1-22}$$

这表示在任何情况下都不收敛。因此，柯西分布随机变量的方差是无穷大。

随机变量函数的数学期望　已知 $Z = g(X, Y)$ 是两个随机变量的函数，Z 的数学期望计算如下：

$$E[Z]=\int_{-\infty}^{\infty}zf_Z(z)\mathrm{d}z$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{4.1-23}$$

为了证明式(4.1-23)可以用来计算 $E[Z]$，我们需要用到一个自变量，这与建立式(4.1-9)类

似。实际上，可以按照式(4.1-10)的思路来进行证明

$$\{z_j < Z \leqslant z_j + \Delta z\} = \bigcup_{k=1}^{N_j} \{(X, Y) \in D_k\}$$

其中 D_k 是包含点 $(x_k^{(j)}, y_k^{(j)})$ 的一系列不相交的小区域，即 $g(x_k^{(j)}, y_k^{(j)}) = z_j$。同时计算两边的概率，且利用 D_k 是不相交的，可得

$$f_Z(z_j)\Delta z_j \approx \sum_{k=1}^{N_j} f(x_k^{(j)}, y_k^{(j)})\Delta a_k^{(j)}$$

其中 $\Delta a_k^{(j)}$ 表示一个无限小的区域。

等号两边同时乘以 z_j，且利用 $z_j = g(x_k^{(j)}, y_k^{(j)})$ 可得

$$z_j f_Z(z_j)\Delta z_j \approx \sum_{k=1}^{N_i} g(x_k^{(j)}, y_k^{(j)}) f_{XY}(x_k^{(j)}, y_k^{(j)})\Delta a_k^{(j)}$$

当 $j \to \infty$，$\Delta z_j \to 0$，$\Delta a_k^{(j)} \to \mathrm{d}a = \mathrm{d}x\mathrm{d}y$ 时，可得

$$\int_{-\infty}^{\infty} z f_Z(z)\mathrm{d}z = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

这里也给出另外一种证明方法①。设 $Z = g(X, Y)$，可得

$$\begin{aligned} E[Z] &= \int_{-\infty}^{\infty} z f_Z(z)\mathrm{d}z \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z f_{Z|Y}(z|y) f_Y(y)\mathrm{d}y\,\mathrm{d}z \end{aligned}$$

根据边缘概率密度函数的定义可以写出上面第二行的表达式。再次利用 $Z = g(X)$ 可得

$$\int_{-\infty}^{\infty} z f_Z(z)\mathrm{d}z = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$$

我们可以利用上面的结论来进行证明。如果固定 Y 的取值为 $Y = y$，函数 $g(X, y)$ 只跟 X 的取值有关，那么 $Y = y$ 时 z 的条件数学期望为

$$\int_{-\infty}^{\infty} z f_{Z|Y}(z|y)\mathrm{d}z = \int_{-\infty}^{\infty} g(x, y) f_{X|Y}(x|y)\,\mathrm{d}x$$

利用上面得到的结论，可以推导出

$$\begin{aligned} E[Z] &= \int_{-\infty}^{\infty} z f_Z(z)\mathrm{d}z \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} z f_{Z|Y}(z|y)\mathrm{d}z\right) f_Y(y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{X|Y}(x|y) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\,\mathrm{d}x\,\mathrm{d}y \end{aligned}$$

证毕。

---

① Carl W. Helstrom, *Probability and Stochastic Processes for Engineers*, 2nd edition. New York, Macmillan, 1991.

例 4.1-9　(独立随机变量乘积的数学期望)设 $g(x,y)=xy$，计算 $E[Z]$，其中 $Z=g(X,Y)$，X 和 Y 服从正态分布且相互独立

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}\left((x-\mu_a)^2+(y-\mu_b)^2\right)\right]$$

解　可将双重积分写为两个积分相乘的形式，直接代入式(4.1-23)可得

$$\begin{aligned}E[Z]&=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}x\exp\left[-\frac{1}{2\sigma^2}(x-\mu_a)^2\right]\mathrm{d}x\\&\quad\times\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}y\exp\left[-\frac{1}{2\sigma^2}(y-\mu_b)^2\right]\mathrm{d}y\\&=\mu_a\mu_b\end{aligned}$$

式(4.1-23)可以用来计算 $E[X]$ 或者 $E[Y]$。对于 $Z=g(X,Y)=X$，可得

$$\begin{aligned}E[X]&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y\\&=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}f_{XY}(x,y)\mathrm{d}y\right]x\,\mathrm{d}x\end{aligned}\tag{4.1-24}$$

根据式(2.6-47)，方括号里面的积分表达式其实是边缘密度函数 $f_X(x)$。因此，式(4.1-23)与下面的定义是完全一致的

$$E[X]\triangleq\int_{-\infty}^{\infty}xf_X(x)\,\mathrm{d}x$$

根据边缘密度的定义可得

$$\begin{aligned}E[X+Y]&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f_{XY}(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{\infty}x\left(\int_{-\infty}^{\infty}f_{XY}(x,y)\mathrm{d}y\right)\mathrm{d}x+\int_{-\infty}^{\infty}y\left(\int_{-\infty}^{\infty}f_{XY}(x,y)\mathrm{d}x\right)\mathrm{d}y\\&=E[X]+E[Y]\end{aligned}\tag{4.1-25}$$

式(4.1-24)可以拓展到 N 个随机变量 X_1，X_2，…，X_N，也就是

$$E\left[\sum_{i=1}^{N}X_i\right]=\sum_{i=1}^{N}E[X_i]\tag{4.1-26}$$

注意到这里并不要求随机变量是相互独立的。

例 4.1-10　(独立的高斯随机变量)设 X，Y 为独立的高斯随机变量，且联合概率密度函数为

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{1}{2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right)$$

因为 X 和 Y 是相互独立的，因此 $f_{XY}(x,y)=f_X(x)f_Y(y)$。利用式(2.6-44)和式(2.6-47)可计算出边缘概率密度函数

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right]$$

$$f_Y(y)=\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left[-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

利用式(4.1-24)可得

$$E[X+Y]=\mu_1+\mu_2$$

例 4.1-11 (Chi 平方定律)在许多科学和工程实践中，信号相加具有无序性，意思是信号之和的功率与各个分信号的功率之和是相同的。比如在光学中，一个表面被不同波长的光源照射，对该表面接收到的功率进行测量，测量结果与各个分光源的功率之和是相同的。在电路中，当电源是由不同频率的正弦波组成，在任何一个电阻上耗散的功率等于各个分频率的功率之和。假设在一个特定的时刻，用独立同分布的高斯随机变量来对这些相互独立的信号源建模。用 $X_1, X_2, \cdots, X_n$ 来表示由 n 个信号源产生的 n 个独立信号。且 X_i: $N(0,1)$，$i=1,2,\cdots,n$，设 $Y_i=X_i^2$。根据第 3 章的式(3.2-2)可知，Y_i 的概率密度函数为

$$f_{Y_i}(y)=\frac{1}{\sqrt{2\pi y}}\mathrm{e}^{-y/2}u(y)$$

对于 $Z_2=Y_1+Y_2$，$Z_3=Y_1+Y_2+Y_3$，$\cdots$，$Z_n=\sum_{i=1}^{n}Y_i$，可通过卷积来轻松计算 Z_2 的概率密度函数

$$\begin{aligned}f_{Z_2}(z)&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi x}}\mathrm{e}^{-x/2}u(x)\times\frac{1}{\sqrt{2\pi(z-x)}}\mathrm{e}^{-\frac{1}{2}(z-x)}u(z-x)\,\mathrm{d}x\\&=\frac{1}{\pi}\mathrm{e}^{-z/2}\int_0^{\sqrt{z}}\frac{\mathrm{d}y}{\sqrt{z-y^2}}u(z)\\&=\frac{1}{2}\mathrm{e}^{-z/2}u(z)\ (\textbf{指数}\ \mathrm{pdf})\end{aligned}$$

令 $x=y^2$ 可以推导出上式的第 1 行到第 2 行。对于第 2 行到第 3 行，这其实就是一个三角函数的积分。为计算 Z_3 的概率密度函数，我们只需将 Z_2 和 Y_3 的概率密度函数卷积即可。

$$\begin{aligned}f_{Z_3}(z)&=\frac{1}{2}\int_{-\infty}^{\infty}\mathrm{e}^{-x/2}u(x)\times\frac{1}{\sqrt{2\pi(z-x)}}\mathrm{e}^{-\frac{1}{2}(z-x)}u(z-x)\,\mathrm{d}x\\&=\frac{1}{\sqrt{2\pi}}z^{\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}z}u(z)\end{aligned}$$

我们将推导的中间过程留给读者，这只涉及一些基本的变换。采用相同的方法，或者数学归纳法，我们可以求出一般表达式如下：

$$f_{Z_n}(z)=\frac{1}{2^{n/2}\Gamma(n/2)}z^{\frac{n-2}{2}}\mathrm{e}^{-z/2}u(z)$$

在第 2 章介绍过这个概率密度函数，当时称为 Chi 平方分布的概率密度函数。更准确的说法应该是“具有 n 个自由度的 Chi 平方分布的概率密度函数”。当 $n>2$ 时，概率密度函数在 $z=0$ 处取值为 0，然后到达顶点，随后呈现为一个单调递减的“尾巴”。当 n 取值较大时，变得非常相似于高斯概率密度函数，而且均值约等于 n。但是，Chi 平方分布是不可能完全变为高斯分布的，因为 Chi 分布的随机变量不可能取负值。图 4.1-2 给出了 n 取不同值时 Chi 平方分布的一些概率密度函数。

可根据定义 $Z_n\triangleq\sum_{i=1}^{n}X_i^2$ 来计算 Chi 平方分布随机变量的均值和方差。其中 $E[Z_n]=$

$E[\sum_{i=1}^{n}X_i^2] = \sum_{i=1}^{n}E[X_i^2] = n$。同样地，$\mathrm{Var}(Z_n) = E[(Z_n - n)^2]$。经过化简可得 $\mathrm{Var}(Z_n) = E[Z_n^2] - n^2$。根据 $E[Z_n^2] = 2n + n^2$，因此有 $\mathrm{Var}(Z_n) = 2n$，最后的推导留给读者。

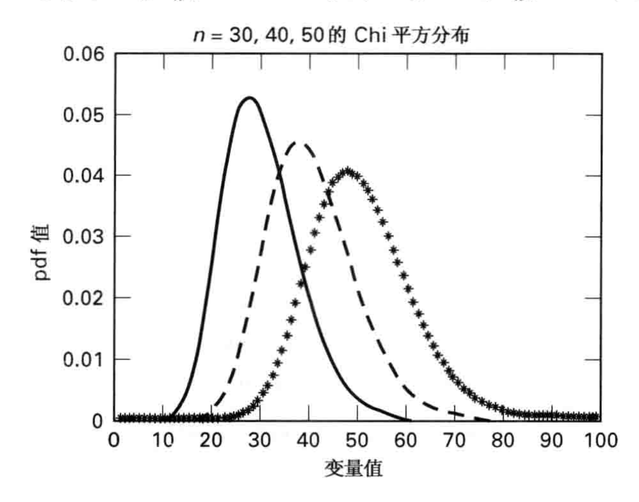


图 4.1-2　$n=30$(实线)，$n=40$(虚线)，$n=50$(星形线)时的 Chi 分布概率密度函数图形。当 n 取值较大时，Chi分布的概率密度函数可用一个服从$N(n,2n)$的正态分布曲线来近似，而且在均值附近取值的概率非常大。例如，对于$n=30$，可假设$X:N(30,60)$，$P[\mu-\sigma<X<\mu+\sigma]=0.6827$。如果采用单精度算法，根据Chi平方分布的概率密度直接计算出来的结果是0.6892

例 4.1-12　在著名的 P 大学，校方要求每个教授都配备一个电子花名册(Rolodex)，这个花名册可以存储班上所有学生的名字。如果教授想抽点一个学生，她只需要点击这个花名册上的“呼叫”按钮，这个时候花名册中的电路就会随机选择一个学生的名字出来。通过使用设备，教授就可以完全不受影响地选择学生出来回答她的问题。用 r 来表示平均呼叫的次数，请问 r 是多少时才能保证每个学生至少被呼叫到一次？

解　使用电子花名册，意味着有的学生一学期下来可能都不会被呼叫到一次，而有的学生可能会被连续呼叫到两次或三次。这取决于这个班级到底有多大。不过平均而言，一个或多个学生没有被呼叫到的情况是非常罕见的。细心的读者可能会注意到：如果我们把呼叫看成小球，把学生看成球洞，这其实就是一个球进洞的问题。设 $R \in \{n, n+1, n+2, \cdots\}$ 表示一次性把 n 个洞都填满的球的个数。这件事能发生的可能就是：前面 $R-1$ 个球把 $n-1$ 个洞填满(事件 E_1)，第 R 个球把剩余的那个洞填满(事件 E_2)。对应到班级点名，意味着经过 $R-1$ 次呼叫后，除了一名学生以外都被呼叫到了(事件 E_1)，然后剩下的这个学生将在第 R 次呼叫中被抽中(事件 E_2)。$P[R=r, n] \triangleq P_R(r, n) = P[E_1E_2] = P[E_1]P[E_2]$，其中 E_1 和 E_2 是独立的。此时有 $P[E_2] = 1/n$，这表示事先指定的一个球进入一个特定洞的概率，根据式(1.8-13) $P[E_1]$ 表示 $P_1(r-1, n)$，即

$$P_1(r-1,n) = \binom{n}{1}\sum_{i=0}^{n-1}\binom{n}{i}(-1)^i\left(1-\frac{i+1}{n}\right)^{r-1}, \quad r \geqslant n$$
$$= 0, \text{ 其他}$$

$P_R(r, n)$ 的表达式为

$$P_R(r,n) = \sum_{i=0}^{n-1}\binom{n}{i}(-1)^i\left(1-\frac{i+1}{n}\right)^{r-1}, \quad r \geqslant n \tag{4.1-27}$$
$$= 0, \text{ 其他}$$

概率值 $P_K(k, n)$ 表示：k 个球进洞后(k 个学生被呼叫到)，所有的 n 个洞(学生)都被填满(呼叫到)的概率，根据式(1.8-9)有

$$P_K(k,n)=\sum_{i=0}^{n}\binom{n}{i}(-1)^i\left(1-\frac{i}{n}\right)^k,\quad k\geqslant n \qquad (4.1\text{-}28)$$
$$=0,\ \text{其他}$$

最后，给出随机变量 R 的数学期望表达式

$$E[R]=\sum_{r=n}^{\infty}r\left(\sum_{i=0}^{n-1}\binom{n}{i}(-1)^i\left(1-\frac{i+1}{n}\right)^{r-1}\right) \qquad (4.1\text{-}29)$$

例 4.1-13 编写一个 MATLAB 程序求解例 4.1-12，计算通过电子花名册点名，保证每个学生都至少被抽中一次需要的呼叫次数 r。假设这个班上有 20 个学生。

解 式(4.1-28)需要通过编程来实现。结果如图 4.1-3 所示。

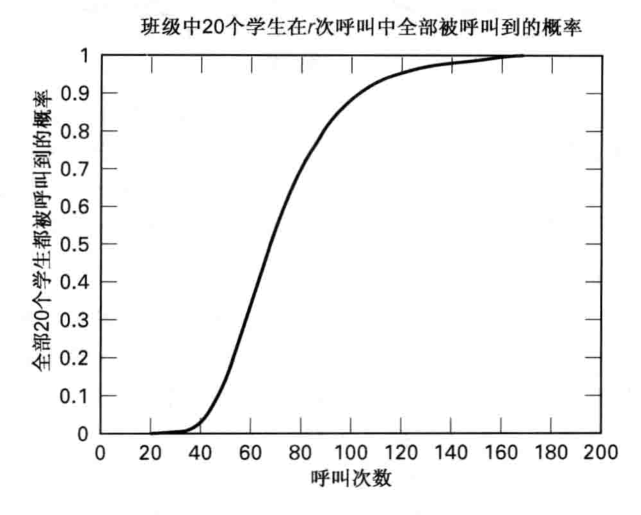


图 4.1-3 例 4.1-13 的 MATLAB 结果

```
function [tries,prob]=occupancy(balls,cells)
tries=1:balls; % identifies a vector ''tries''
prob=zeros(1,balls); % identifies a vector ''prob''
a=zeros(1,cells); % identifies a vector ''a''
d=zeros(1,cells); % identifies a vector ''d''

term=zeros(1,cells); % identifies a vector ''term''
% next follows the realization of Equation (4.1-27)
for m=1:balls

    for k=1:cells
        a(k)=(-1)^k)*prod(1:cells)/(prod(1:k)*prod(1:cells-k));
        d(k)=(1-(k/cells))^m;
        term(k)=a(k)*d(k);
    end
prob(m)=1+sum(term);
end
```

```
  plot(tries,prob)
  title(['Probability of all ' num2str(cells) ' students in the class
being called in r tries'])
xlabel('number of tries')
ylabel(['Probability of all ' num2str(cells) ' students being called'])
```

例 4.1-14　编写一个 MATLAB 程序来计算平均需要呼叫多少次，才能保证每个学生至少被抽中一次。假设最多有 50 个学生，而且要确保呼叫的次数足够大 ($n \geqslant 400$)。

解　式(4.1-29)需要通过编程来实现。结果如图 4.1-4 所示。

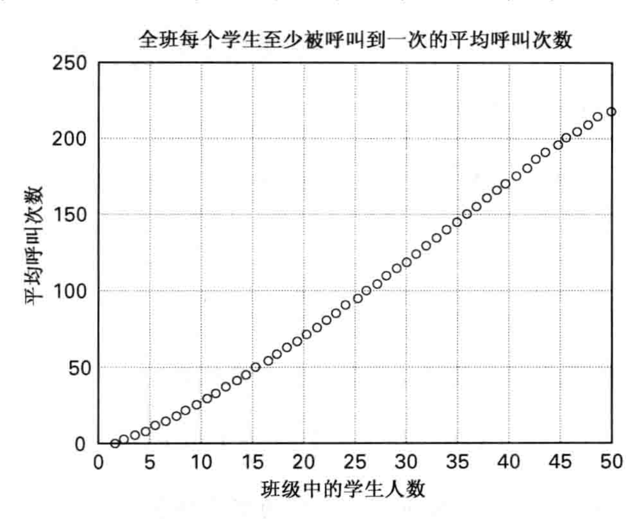


图 4.1-4　例 4.1-14 的 MATLAB 结果

```
function [cellvec,avevec]=avertries(ballimit,cellimit);
cellvec = 1:cellimit;
termvec = zeros(1,ballimit);
avevec = zeros(1,cellimit);
brterm=zeros(1,ballimit);
srterm=zeros(1,ballimit);
for n=1:cellimit;
   a = zeros(1,n);
   d = zeros(1,n);
   termvec = zeros(1,n);
   for r=1:ballimit
      for i=1:n-1
         a(i) = ((-1)^i)*prod(1:n-1)/(prod(1:i)*prod(1:n-1-i));
         d(i) = (1-((i-1)/n))^(r-1);
         termvec(i) = a(i)*d(i);
      end
      brterm(r)=r*sum(termvec);
      lrterm(r)=r*((1-(1/n)))^(r-1);
   end
   avevec(n)=sum(brterm)+sum(lrterm);
end
plot(cellvec,avec,'o')
```

```
title('Average number of Rolodex tries to call all students at least
once')
xlabel('number of students in the class')
ylabel('Expected number of Rolodex tries to reach all students at least
once')
grid
```

例 4.1-15 (几何分布)随机变量 X 服从几何分布，其概率质量函数为

$$P_X(n) = (1-a)a^n u(n)$$

其中 u 表示单位阶跃函数①，且 $0 < a < 1$。根据等比数列求和 $\sum_{n=0}^{\infty} a^n = (1-a)^{-1}, 0 < a < 1$，易知 $\sum_{n=0}^{\infty} P_X(n) = 1$。数学期望为

$$E[X] = \mu = (1-a)\sum_{n=0}^{\infty} na^n = (1-a) \times a \times \frac{\mathrm{d}}{\mathrm{d}a}\{(1-a)^{-1}\} = \frac{a}{1-a}$$

求出 a 的表达式

$$a = \frac{\mu}{1+\mu}$$

此时，几何分布的概率质量函数可以重写为如下表达式：

$$P_X(n) = \frac{1}{1+\mu}\left(\frac{\mu}{1+\mu}\right)^n u(n)$$

注意：另外还有一种几何分布随机变量的定义，它的概率质量函数取值范围是 $[1, \infty)$ 而不是 $[0, \infty)$。在例 1.9-4 中出现过这种几何分布的定义，在那里它的概率质量函数表达式为 $P_X(n) = (1-a)a^{n-1}u(n-1)$，这其实和此处的定义是一致的，只不过序列右移了一位而已。

4.2 条件期望

在许多情况下，我们希望知道总体中一个子集的平均值。例如，一次考试中及格试卷的平均分；年龄为 70 岁的正常人平均预期寿命；战斗机飞行员的平均体重(许多空军对飞行员的可接受体重都规定了一个上限和下限)；长跑运动员的平均血压，等等。这类问题都可以归纳到条件期望的范畴。

在条件期望中，我们计算总体中一个子集的平均值，而且由于一个特定事件的输出值使得它们都具有共同的性质。比如要计算及格试卷的平均分，这里的子集就是那些及格的试卷。那么这些试卷的共同特性是什么？那就是它们的分数，比如都大于等于 65 分。而发生的事件就是指它们都及格了。

定义 4.2-1 事件 B 发生时，X 的条件期望定义如下：

$$E[X|B] \triangleq \int_{-\infty}^{\infty} x f_{X|B}(x|B)\,\mathrm{d}x \qquad (4.2\text{-}1)$$

如果 X 是离散的，式(4.2-1)替换为

① 当 $n \geqslant 0$ 时 $u(n) = 1$，其余时刻 $u(n) = 0$。

$$E[X|B] \triangleq \sum_i x_i P_{X|B}(x_i|B) \tag{4.2-2}$$

为了给读者关于条件期望一个直观的认识，考虑一次概率论考试中的分数：28，35，44，66，68，75，77，80，85，87，90，100，100。假设及格分数线为65 分。在这次考试中平均分是71.9，但是平均及格分数为82.8。下面给出一个联系非常紧密的例题。

例4.2-1　（均匀分布的条件期望）考虑一个连续随机变量 X 和一个事件 $B \triangleq \{X \geqslant a\}$。利用式(2.6-1)和式(2.6-2)可得

$$F_{X|B}(x|X \geqslant a) = \begin{cases} 0, & x < a \\ \dfrac{F_X(x) - F_X(a)}{1 - F_X(a)}, & x \geqslant a \end{cases} \tag{4.2-3}$$

因此有

$$f_{X|B}(x|X \geqslant a) = \begin{cases} 0, & x < a \\ \dfrac{f_X(x)}{1 - F_X(a)}, & x \geqslant a \end{cases} \tag{4.2-4}$$

以及

$$E[X|X \geqslant a] = \frac{\displaystyle\int_a^\infty x f_X(x)\,\mathrm{d}x}{\displaystyle\int_a^\infty f_X(x)\,\mathrm{d}x} \tag{4.2-5}$$

假设 X 是区间[0，100]上的均匀分布随机变量，可得

$$E[X] = \frac{1}{100}\int_0^{100} x\ \mathrm{d}x = 50$$

如果设 $a = 65$，根据式(4.2-5)可得

$$E[X|X \geqslant 65] = 82.5$$

当随机变量是相互联系的时候，就会经常用到条件期望。例如用 Y 来表示随机选取的一个人的寿命，设 X 是一个二元随机变量，用来区分这个人吸烟与否。当 $X = 0$ 时表示这个人不吸烟，$X = 1$ 表示这个人吸烟。很显然，一般会认为 $E[Y| X = 0]$ 比 $E[Y| X = 1]$ 大①。或者假设 X 表示入射光的照射强度，Y 表示光电探测器产生的瞬时光电流强度。一般而言，光照越强 Y 的数学期望会越大，反之越小。下面我们将给出一些重要的概念。

定义4.2-2　设 X 和 Y 是离散随机变量，联合概率质量函数为 $P_{X,Y}(x_i, y_j)$。用 $E[Y|X = x_i]$ 来表示给定 $X = x_i$ 时 Y 的条件期望

$$E[Y|X = x_i] \triangleq \sum_j y_j P_{Y|X}(y_j|x_i) \tag{4.2-6}$$

在这里 $P_{Y|X}(y_j | x_i)$ 是一个条件概率表达式，表示事件 $\{X = x_i\}$ 发生的前提下，事件 $\{Y = y_j\}$ 发生的概率，其值为 $P_{X,Y}(x_i, y_j)/P_X(x_i)$。

下面我们推导出一个非常有用的公式，可以用 Y 在 $X = x$ 前提时的条件期望来计算 $E[Y]$。这么做的原因类似于总体概率的计算，即可以通过条件概率来计算一个事件的平均或总体概率[参见式(1.6-7)或式(2.6-4)]。因此可得

① 统计表明：每抽一支烟寿命大约减少8 分钟，因此如果每天抽一包烟，一年下来寿命将会减少40 天！

$$E[Y] = \sum_j y_j P_Y(y_j) \tag{4.2-7}$$

$$= \sum_j y_j \sum_i P_{X,Y}(x_i, y_j)$$

$$= \sum_i \left[\sum_j y_j P_{Y|X}(y_j|x_i) \right] P_X(x_i)$$

$$= \sum_i E[Y|X = x_i] P_X(x_i) \tag{4.2-8}$$

式(4.2-8)是一个非常简洁的结论，意味着对 Y 在 X 的前提条件下所有可能的条件期望①取平均就可以得到 $E[Y]$。因此，对于先前的那个吸烟者寿命的问题，假设 $E[Y|X=0]=79.2$（单位：岁），$E[Y|X=1]=69.4$（单位：岁），而且 $P_X(0)=0.75$，$P_X(1)=0.25$。那么可计算出普通人群的预期寿命如下：

$$E[Y] = 79.2 \times 0.75 + 69.4 \times 0.25 = 76.75$$

对于连续时间的情况，也有一个与式(4.2-8)类似的结果。根据第 2 章中的式(2.6-85)可得

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}, \qquad f_X(x) \neq 0 \tag{4.2-9}$$

下面给出连续随机变量条件期望的定义。

定义 4.2-3 设 X 和 Y 是连续随机变量，联合概率密度函数为 $f_{XY}(x, y)$。当 $X=x$ 时，用式(4.2-9)来表示 Y 的条件概率密度函数。那么当 $X=x$ 时，Y 的条件期望为

$$E[Y|X = x] \triangleq \int_{-\infty}^{\infty} y f_{Y|X}(y|x) \mathrm{d}y \tag{4.2-10}$$

根据下面的公式

$$E[Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{XY}(x,y) \, \mathrm{d}x \, \mathrm{d}y \tag{4.2-11}$$

由式(4.2-9)和式(4.2-10)可得

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} f_X(x) \left[\int_{-\infty}^{\infty} y f_{Y|X}(y|x) \mathrm{d}y \right] \mathrm{d}x \\ &= \int_{-\infty}^{\infty} E[Y|X = x] f_X(x) \, \mathrm{d}x \end{aligned} \tag{4.2-12}$$

与式(4.2-8)等价，式(4.2-12)是一个连续随机变量。用来计算 $E[Y]$ 的时候具有很大的优势(用直接法)。我们用光通信的一个例子来讲解这个观点。

例 4.2-2 （条件泊松分布）对于图 4.2-1 所示的光电探测器，在时间 τ 内产生的光电子数目 Y 取决于(归一化)入射光能量 X。如果 X 是常数(即 $X=x$)，那么 Y 是一个服从参数为 x 的泊松随机变量[4-4]。而实际光源(稳定的激光器除外)无法发射能量恒定的信号，因此 X 应该被看成一个随机变量。在某些特定情况下，X 的概率密度函数可以建模为

① 这意味着我们把 Y 关于 X 的条件期望看成一个随机数。接下来将会详细讲解这个重要概念。现在我们认为 X 为一个给定的值 x_i(或 x)。

$$f_X(x)=\begin{cases}\dfrac{1}{\mu_X}\exp\left(-\dfrac{x}{\mu_X}\right), & x\geqslant 0\\ 0, & x<0\end{cases} \tag{4.2-13}$$

其中参数μ_X表示$E[X]$。在此，将根据式(4.2-12)用直接法来计算$E[Y]$。

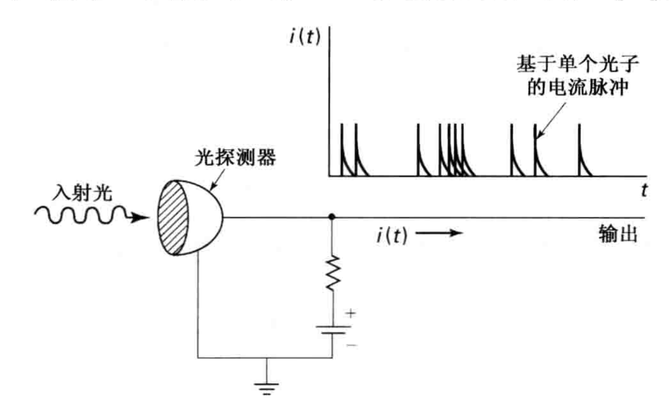


图4.2-1 在光电探测器中，入射光产生一个由光生电子组成的电流

解 因为当$X=x$时Y服从泊松分布，我们可以写出

$$P[Y=k|X=x]=\frac{x^k}{k!}\mathrm{e}^{-x}, \qquad k=0,1,2,\cdots$$

再根据例4.1-6可得

$$E[Y|X=x]=x$$

最后，利用式(4.2-12)的结论，通过合理的变量替换可得

$$E[Y]=\int_0^{\infty}x\left[\frac{1}{\mu_X}\exp\left(-\frac{x}{\mu_X}\right)\right]\mathrm{d}x$$

通过分部积分可得

$$E[Y]=\mu_X$$

为了体现出可以使用一个简易的方法来求出结果，在此采用直接法计算

$$E[Y]=\sum_{k=0}^{\infty}kP_Y(k) \tag{4.2-14}$$

为计算$P_Y(k)$，可对式(4.2-13)的$f_X(x)$进行变换[参见式(2.6-14)]，最终的结果为[参见式(2.6-23)]

$$P_Y(k)=\frac{\mu_X^k}{(1+\mu_X)^{k+1}} \tag{4.2-15}$$

最后，将式(4.2-15)代入式(4.2-14)可得

$$E[Y]=\sum_{k=0}^{\infty}k\frac{\mu_X^k}{(1+\mu_X)^{k+1}}$$

易知上面序列的求和结果为μ_X。还可以采用其他巧妙的方法来间接计算这个序列之和，比如利用导数。

例4.2-3 (条件高斯分布)设X和Y是零均值的随机变量，联合概率密度函数为

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}}\exp\left(-\frac{x^2+y^2-2\rho xy}{2\sigma^2(1-\rho^2)}\right),\quad |\rho|<1 \tag{4.2-16}$$

我们很快会发现(参见4.3节),式(4.2-16)其实是两个随机变量的联合高斯概率密度函数的一种特例。首先可以看出,当$\rho\neq0$时$f_{XY}(x,y)\neq f_X(x)f_Y(y)$,因此$\rho\neq0$时$X$和$Y$不是相互独立的。当$\rho=0$时有$f_{XY}(x,y)=f_X(x)f_Y(y)$,所以$\rho=0$意味着$X$和$Y$独立。我们暂时对参数$\rho$($\rho$称为归一化协方差或协相关系数)不熟悉,只要知道当$\rho=0$时,$X$和$Y$是零均值的高斯随机变量

$$f_X(x)=f_Y(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}$$

不过,即使Y是一个零均值的随机变量,$X=x$时Y的条件期望并不为0。因为根据式(4.2-9)

$$f_{Y|X}(y|x)=\frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}}\exp\left(-\frac{(y-\rho x)^2}{2\sigma^2(1-\rho^2)}\right) \tag{4.2-17}$$

可知,$f_{Y|X}(y|x)$是均值为ρx的高斯密度函数,即

$$\begin{aligned}E[Y|X=x]&=\int_{-\infty}^{\infty}yf_{Y|X}(y|x)\,dy\\&=\rho x\end{aligned} \tag{4.2-18}$$

当ρ趋向于1时,$E[Y|X=x]\approx x$,此时意味着Y非常接近于X(如果$\rho=1$意味着X和Y完全一致)。如果我们希望估计Y的取值,用Y_P来表示。而且此时观测值$X=x$,那么最佳的估计策略就是设$Y_P=x$。另一方面,如果$\rho=0$,意味着此时X的观测值对于Y的估计毫无帮助。从高斯分布的这个例子可以看出,参数ρ表征了通过观测一个随机变量预测另外一个随机变量的能力。在此需要注意的是:一个随机变量不能线性预测另外一个随机变量,这并不意味着二者是相互独立的。

例4.2-4 (随机变量之和的条件期望)设K_1和K_2是两个独立的离散随机变量,试计算条件期望$E[K_1|K_1+K_2=m]$。首先需要计算条件概率$P[K_1=k_1|K_1+K_2=m]$。这个条件概率可计算如下:

$$\begin{aligned}P[K_1=k_1|K_1+K_2=m]&=\frac{P[K_1=k_1,K_1+K_2=m]}{P[K_1+K_2=m]}\\&=\frac{P[K_1=k_1,K_2=m-k_1]}{P[K_1+K_2=m]}\\&=\frac{P[K_1=k_1]P[K_2=m-k_1]}{P[K_1+K_2=m]}\end{aligned} \tag{4.2-19}$$

在此设K_1和K_2都服从参数为θ的泊松分布,因此这两个随机变量是独立同分布的。通过式(4.2-19)以及例3.3-8的结论可知

$$\begin{aligned}&P[K_1=k_1|K_1+K_2=m]\\&=\frac{(e^{-\theta_1}\theta_1^{k_1}/k_1!)\times(e^{-\theta_2}\theta_2^{m-k_1}/(m-k_1)!)}{e^{-(\theta_1+\theta_2)}(\theta_1+\theta_2)^m/m!}\\&=\binom{m}{k_1}\theta_1^{k_1}\theta_2^{m-k_1}\times(\theta_1+\theta_2)^{-m}\end{aligned} \tag{4.2-20}$$

再次根据 $E[K_1 \mid K_1 + K_2 = m] \triangleq \sum_{k_1=0}^{m} k_1 P[K_1 = k_1 \mid K_1 + K_2 = m]$，以及二项式展开公式 $\sum_{k=0}^{n} \binom{n}{k} \theta_1^k \theta_2^{n-k} = (\theta_1 + \theta_2)^n$。根据式(4.2-20)最终可得

$$E[K_1|K_1 + K_2 = m] = m \times \left(\frac{\theta_1}{\theta_1 + \theta_2}\right) \tag{4.2-21}$$

例 4.2-5 （续例 4.2-4）设 K_1，K_2，K_3 代表 $l = 3$ 的多项式随机变量，也就是存在三个可能输出。对于 n 次试验，其概率质量函数为①

$$\begin{aligned} P_{\boldsymbol{K}}(k_1, k_2, k_3) &= P[K_1 = k_1, K_2 = k_2, K_3 = k_3] \\ &= \begin{cases} \frac{n!}{k_1!k_2!k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}, & k_1 + k_2 + k_3 = n, \text{所有 } k_i \geqslant 0 \\ 0, & \text{其他} \end{cases} \end{aligned} \tag{4.2-22}$$

其中 $p_1 + p_2 + p_3 = 1$，在此我们希望计算 $E[K_1 \mid K_1 + K_2 = m]$。

解 同上一例题，需要首先计算 $P[K_1 = k_1 \mid K_1 + K_2 = m]$，过程如下：

$$P[K_1 = k_1|K_1 + K_2 = m] = \frac{P[K_1 = k_1, K_1 + K_2 = m]}{P[K_1 + K_2 = m]}$$

对于多项式的情况，事件 $\{\zeta: K_1(\zeta) + K_2(\zeta) = m\} \cap \{\zeta: K_1(\zeta) = k_1\}$ 与事件 $\{\zeta: K_1(\zeta) = k_1, K_2(\zeta) = m - k_1, K_3(\zeta) = n - m\}$ 是等价的，因此有

$$P[K_1 = k_1|K_1 + K_2 = m] = \frac{P[K_1 = k_1, K_2 = m - k_1, K_3 = n - m]}{P[K_3 = n - m]} \tag{4.2-23}$$

$$\begin{aligned} &= \frac{n!}{k_1!(m - k_1)!(n - m)!} p_1^{k_1} p_2^{m-k_1} p_3^{n-m} \\ &\quad \div \frac{n!}{(n - m)!m!} p_3^{n-m} (1 - p_3)^m \end{aligned}$$

$$= \binom{m}{k_1} p_1^{k_1} p_2^{m-k_1} (p_1 + p_2)^{-m} \tag{4.2-24}$$

根据公式

$$E[K_1|K_1 + K_2 = m] = \sum_{k_1} k_1 P[K_1 = k_1|K_1 + K_2 = m]$$

可得

$$E[K_1|K_1 + K_2 = m] = m \frac{p_1}{p_1 + p_2} \tag{4.2-25}$$

下面的这个结论留给读者来完成

$$E[K_2|K_1 + K_2 = m] = m \frac{p_2}{p_1 + p_2} \tag{4.2-26}$$

这种类型的问题在估计过程中经常会遇到，即所谓的期望最大化算法，我们将在第 11 章详细讨论这个问题。

---

① 这里表达方式与二项分布的不太一样。如果要把这里的多项式分布的记号套用到二项分布的情况，那么就需要把二项分布随机变量 K 设为 $K_1 = K$ 和 $K_2 = n - K$。对于一般情况下的 l 项多项式分布，我们需要时刻保证 $K_1 + K_2 + \cdots + K_l = n$。

条件期望被当成一个随机变量

考虑函数 $Y = g(X)$ 这个特例，其中 X 是一个离散随机变量，那么 Y 的数学期望为

$$\begin{aligned} E[Y] &= \sum_i g(x_i)P_X(x_i) \\ &= E[g(X)] \end{aligned}$$

这意味着可以把式(4.2-8)写为一个与之相似的形式，即

$$\begin{aligned} E[Y] &= \sum_i E[Y|X = x_i]P_X(x_i) \\ &= E[E[Y|X]] \end{aligned} \tag{4.2-27}$$

需要注意的是：$E[Y|X = x_i]$ 是一个数值，类似于函数 $g(x_i)$ 的值，而 $E[Y|X]$ 就是随机变量 X 的一个函数，而且 $E[Y|X]$ 本身也是一个随机变量。给定一个概率空间 $\mathscr{P} = (\Omega, \mathscr{F}, P)$ 以及定义在 $\mathscr{P}$ 上面的随机变量 X，对于每个输出值 $\zeta \in \Omega$ 我们产生对应的一个实数 $E[Y|X = X(\zeta)]$。当把变量 ζ 当成基本实验的输出时，对应着一个取值 $E[Y|X = X(\zeta)]$，因此可以把 $E[Y|X]$ 看成一个随机变量。一般而言，X 关于 ζ 的函数特性往往被忽视了，而且我们也很少在基本概率空间 $\mathscr{P}$ 上描述 X。下面的例子将讲解如何把条件期望当成一个随机变量来使用。

例 4.2-6 (多通道通信)考虑这么一个通信系统，信息传输的时延是 T(单位：ms)，传输信道的选项为 L。其中 $L=1$ 表示卫星信道，$L=2$ 表示同轴电缆信道，$L=3$ 表示微波表面链接，$L=4$ 表示光纤链接。一般根据传输能力来选择信道，这完全是随机的。假设 $P_L(l) = 1/4$，$l = 1, \cdots, 4$，而且已知 $E[T|L = 1] = 500$，$E[T|L = 2] = 300$，$E[T|L = 3] = 200$ 以及 $E[T|L = 4] = 100$。那么随机变量 $g(L) \triangleq E[T|L]$ 的定义如下：

$$g(L) = \begin{cases} 500, & L = 1 \quad P_L(1) = \frac{1}{4} \\ 300, & L = 2 \quad P_L(2) = \frac{1}{4} \\ 200, & L = 3 \quad P_L(3) = \frac{1}{4} \\ 100, & L = 4 \quad P_L(4) = \frac{1}{4} \end{cases}$$

可得 $E[T] = E[g(L)] = 500 \times \frac{1}{4} + 300 \times \frac{1}{4} + 200 \times \frac{1}{4} + 100 \times \frac{1}{4} = 275$。

$E[Y|X]$ 表示一个随机变量，这不仅适用于离散随机变量，而且对于连续或混合的随机变量也是同样适用的。比如，对于例 4.2-12

$$E[Y] = \int_{-\infty}^{\infty} E[Y|X = x]f_X(x)\,\mathrm{d}x$$

上式也同样可以写成 $E[Y] = E[E[Y|X]]$，这里把 $E[Y|X]$ 看成连续随机变量 X 的一个函数。括号里面的数学期望是对 Y 而言的，而括号外面是对 X 而言的。

上面讲到的内容还可以扩展到更复杂的情况。比如，$E[Z|X, Y]$ 就是随机变量 X 和 Y 的函数，因此这是一个随机变量的二元方程。一个特定的输出 $\zeta \in \Omega$，对应于一个值 $E[Z|X(\zeta), Y(\zeta)]$。为了计算 $E[Z]$，我们需要写成 $E[Z] = E[E[Z|X, Y]]$ 的形式，以连续随机变量为例，$E[Z]$ 的计算过程如下：

$$\begin{aligned} E[Z] &= E[E[Z|X,Y]] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z f_{Z|X,Y}(z|x,y)\, f_{XY}(x,y)\mathrm{d}x\mathrm{d}y\mathrm{d}z \end{aligned} \tag{4.2-28}$$

在本节最后，我们将条件期望的一些特性归纳于此。

条件期望的一些特性

特性(i)　$E[Y]=E[E[Y|X]]$。

证明　输入变量为离散时的证明参见式(4.2-8)，连续的情况参见式(4.2-12)。其中括号最里面是关于 Y 的数学期望，括号外面一层是关于 X 的数学期望。

特性(ii)　如果 X 和 Y 是相互独立的，那么有 $E[Y|X]=E[Y]$。

证明：

$$E[Y|X=x]=\int_{-\infty}^{\infty} y f_{Y|X}(y|x)\mathrm{d}y$$

如果 X 和 Y 是相互独立的，$f_{XY}(x,y)=f_{Y|X}(y|x)f_X(x)=f_Y(y)f_X(x)$ 成立，此时可得 $f_{Y|X}(y|x)=f_Y(y)$，以及

$$E[Y|X=x]=\int_{-\infty}^{\infty} y f_Y(y)\mathrm{d}y=E[Y]$$

上式对于每个 x 都成立，所以有

$$E[Y|X]=\int_{-\infty}^{\infty} y f_Y(y)\mathrm{d}y=E[Y]$$

离散情况的证明与之类似。

特性(iii)　$E[Z|X]=E[E[Z|X,Y]|X]$

证明：

$$\begin{aligned}E[Z|X=x]&=\int_{-\infty}^{\infty} z f_{Z|X}(z|x)\mathrm{d}z\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z f_{Z|X,Y}(z|x,y)f_{Y|X}(y|x)\mathrm{d}z\,\mathrm{d}y\\&=\int_{-\infty}^{\infty}\mathrm{d}y\, f_{Y|X}(y|x)\int_{-\infty}^{\infty} z f_{Z|X,Y}(z|x,y)\mathrm{d}z\\&=E\left[E[Z|X,Y]|X=x\right]\end{aligned}$$

内层的数学期望是对于 Z 而言，而外层是对于 Y 求数学期望。因为上式对于所有的 x 都成立，因此有 $E[Z|X]=E[E[Z|X,Y]|X]$。均值 $\mu_Y=E[Y]$ 是对随机变量 Y 的一个估计，估计的均方误差为 $\varepsilon^2=E[(Y-\mu_Y)^2]$。实际上 μ_Y 是一个最优估计，任何 μ_Y 之外的估计值都会导致 ε^2 增大。

4.3　随机变量的矩

尽管数学期望可以很好地“总结”一个随机变量的特性，但要想完全描述这个随机变量这还远远不够。在 4.1 节中，我们就曾经见到均值相同的两组数据集合，但它们的样本偏差却差异很大。同样地，两个随机变量的均值可能一样，但它们的标准偏差却不同。在此把 μ_X，σ_X^2 和 $E[X^2]$ 等统称为 X 的矩(moment)。一般而言，一个随机变量往往拥有许多非零的高阶矩。在一些特定的条件下(参见 4.5 节)，它们完全可以描述一个随机变量的特性，也可以利用这

些高级矩来重构一个随机变量的概率密度函数。在下面的定义中，假设随机变量的矩都是存在的(不过并不是所有随机变量都存在矩)。

定义 4.3-1 X 的 r 阶矩的定义为

$$m_r \triangleq E[X^r] = \int_{-\infty}^{\infty} x^r f_X(x)\,\mathrm{d}x, \qquad \text{其中 } r = 0, 1, 2, 3, \cdots \tag{4.3-1}$$

如果 X 是一个离散随机变量，那么可以用它的概率质量函数来计算 r 阶矩

$$m_r \triangleq \sum_i x_i^r P_X(x_i)$$

其中 $m_0 = 1$，$m_1 = \mu$ (均值)。

定义 4.3-2 X 的 r 阶中心矩定义如下：

$$c_r \triangleq E[(X-\mu)^r], \qquad \text{其中 } r = 0, 1, 2, 3, \cdots \tag{4.3-2a}$$

对于一个离散随机变量，c_r 可以计算为

$$c_r \triangleq \sum_i (x_i - \mu)^r P_X(x_i) \tag{4.3-2b}$$

c_2 是最常用的中心矩，也称为方差，常用 σ^2 或者 $\mathrm{Var}[X]$ 来表示。记住 $c_0 = 1$，$c_1 = 0$，$c_2 = \sigma^2$。下面有一个重要的公式，把方差 $E[X^2]$ 和 μ 联系在一起

$$\sigma^2 = E\left[[X-\mu]^2\right] = E[X^2] - E[2\mu X] + E[\mu^2]$$

对于任意的常数 a，有 $E[aX] = aE[X]$ 以及 $E[a^2] = a^2$ 成立，因此有

$$\begin{aligned} \sigma^2 &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned} \tag{4.3-3}$$

其中有 $E[X] \triangleq \mu$。为了表达更简洁，常用一个上画线来表示数学期望，即 $\overline{X^r} = E[X^r]$，同样也适用于各阶矩。利用这种表示方法，式(4.3-3)可以写为

$$\sigma^2 = \overline{X^2} - \mu^2 \tag{4.3-4a}$$

或者等价的表达式为

$$\overline{X^2} = \sigma^2 + \mu^2 \tag{4.3-4b}$$

式(4.3-4a)把 μ_2 以及 μ 同二阶中心矩 c_2 联系在一起。我们还可以把这个结论推广到一般情况

$$(X-\mu)^r = \sum_{i=0}^{r} \binom{r}{i} (-1)^i \mu^i X^{r-i} \tag{4.3-5a}$$

对式(4.3-5a)两边同时取数学期望，再根据数学期望的线性特性可得

$$c_r = \sum_{i=0}^{r} \binom{r}{i} (-1)^i \mu^i m_{r-i} \tag{4.3-5b}$$

例 4.3-1 设 X 为二项分布随机变量，试计算 m_2。根据定义

$$P_X(k) = \binom{n}{k} p^k q^{n-k}$$

以及

$$\begin{aligned} m_2 &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} \\ &= p^2 n(n-1) + np \\ &= n^2p^2 + npq \end{aligned} \tag{4.3-6}$$

如何从上面的第 2 行推导第 3 行，这需要用到一些代数方面的知识，我们把这些留给读者作为练习。从第 3 行到第 4 行，需要用到 $q \triangleq 1-p$。因此，X 的数学期望为

$$\begin{aligned} m_1 &= \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k q^{n-k} \\ &= np = \mu \end{aligned} \tag{4.3-7}$$

根据式(4.3-6)的结论以及式(4.3-4)，我们可以归纳出如下的一般结论，其中二项分布随机变量的概率质量函数为 $b(k;\ n,\ p)$

$$\sigma^2 = npq \tag{4.3-8}$$

对于任意给定的 n，当 $p=q=0.5$ 时方差取最大值(参见图 4.3-1)。

例 4.3-2　(零均值高斯随机变量的二阶矩)试计算随机变量 X 的二阶矩 c_2，其中 X：$N(0,\ \sigma^2)$。因为 $\mu=0$，$c_2=m_2$，所以有

$$c_2 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 \mathrm{e}^{-\frac{1}{2}(x/\sigma)^2}\,\mathrm{d}x$$

上面的积分结果可参见例 4.1-7，其中有 $E[X^2]=\sigma^2$。因此，无论高斯随机变量 X 是否是零均值的，其方差都是 σ^2。

下面将用一些有趣的甚至有些难度的例题来讲解矩的应用。

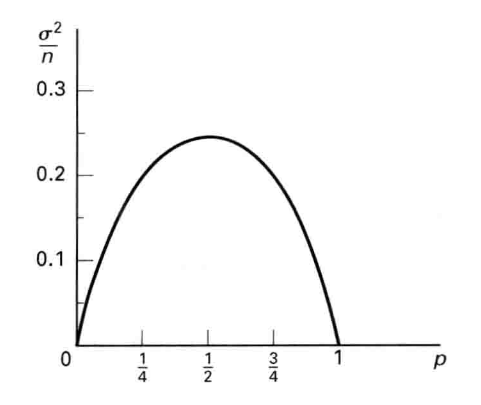


图 4.3-1　二项分布随机变量的方差，参数为 p

例 4.3-3　(熵)如果随机变量 X 的概率密度函数 $f_X(x)$ 未知，我们想用函数 $p(x)$ 来估计 $f_X(x)$，根据最大熵(ME)准则，应该使函数 $p(x)$ 的熵最大，定义为[4-5]

$$H[X] \triangleq -\int_{-\infty}^{\infty} p(x) \ln p(x)\,\mathrm{d}x \tag{4.3-9}$$

同时还应该满足下面的限制条件

$$p(x) \geqslant 0 \tag{4.3-10a}$$

$$\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1 \tag{4.3-10b}$$

$$\int_{-\infty}^{\infty} x p(x)\,\mathrm{d}x = \mu \tag{4.3-10c}$$

$$\int_{-\infty}^{\infty} x^2 p(x)\,\mathrm{d}x = m_2,\ \text{等等} \tag{4.3-10d}$$

假设通过某种测量知道了式(4.3-10c)中的 μ，而且假设 $x \geqslant 0$。因此，我们希望找到一个合适的 $p(x)$ 使得式(4.3-9)中的 $H[X]$ 最大化，当然还要满足式(4.3-10)中前三个限制条件。根

据拉格朗日乘数法[4-6]可知，对 $p(x)$ 求导所求的解可使下面表达式取最大值：

$$-\int_0^{\infty} p(x)\ln p(x)\,\mathrm{d}x-\lambda_1\int_0^{\infty} p(x)\,\mathrm{d}x-\lambda_2\int_0^{\infty} xp(x)\,\mathrm{d}x$$

常数 λ_1 和 λ_2 即拉格朗日乘数。通过求导可得

$$\ln p(x)=-(1+\lambda_1)-\lambda_2 x$$

或者

$$p(x)=\mathrm{e}^{-(1+\lambda_1+\lambda_2 x)} \tag{4.3-11}$$

将这个结果代入式(4.3-10b)和式(4.3-10c)可以得到

$$\mathrm{e}^{-(1+\lambda_1)}=\frac{1}{\mu},\quad \mu>0$$

以及

$$\lambda_2=\frac{1}{\mu}$$

因此，$f_X(x)$ 的最大熵估计值为

$$p(x)=\begin{cases}\dfrac{1}{\mu}\mathrm{e}^{-x/\mu}, & x\geqslant 0\\ 0, & x<0\end{cases} \tag{4.3-12}$$

如果 μ 和 σ^2 都是已知的，如何求解此时 $f_X(x)$ 的最大熵估计值？将这个问题留给读者作为练习。在这种情况下，$p(x)$ 是一个均值 μ，方差为 σ^2 的正态分布。

常见的均值，方差和均方值归纳表　表 4.3-1 归纳了一些常见连续随机变量的均值，方差和均方值。本书已经给出了部分结果的计算过程，剩下的那些留做本章的习题。

表 4.3-2 给出了一些常见的离散随机变量的结果。

绝对矩和一般矩比起 m_r 和 c_r 来，用处偏少，但不失一般性，我们仍然在这里给出它们的定义

$$E[|X|^r]\triangleq\int_{-\infty}^{\infty}|x|^r f_X(x)\,\mathrm{d}x \quad (\text{绝对矩})$$

$$E[(X-a)^r]\triangleq\int_{-\infty}^{\infty}(x-a)^r f_X(x)\,\mathrm{d}x \quad (\text{一般矩})$$

如果设 $a=\mu$，那么关于 a 的一般矩就成为中心矩。如果 $a=0$，一般矩就是 m_r。

表 4.3-1　一些常见连续随机变量的均值，方差和均方值

分布簇	pdf $f(x)$	均值 $\mu=E[X]$	方差 σ^2	均方值 $E[X^2]$
均匀分布	$U(a,b)$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	$\frac{1}{3}(b^2+ab+a^2)$
指数分布	$\frac{1}{\mu}\mathrm{e}^{-x/\mu}u(x)$	μ	μ^2	$2\mu^2$
高斯分布	$\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$\mu^2+\sigma^2$
拉普拉斯分布	$\frac{1}{\sqrt{2}\sigma}\mathrm{e}^{-\frac{\sqrt{2}}{\sigma}\lvert x\rvert}$	0	σ^2	σ^2
瑞利分布	$\frac{x}{\sigma^2}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}u(x)$	$\sqrt{\frac{\pi}{2}}\sigma$	$\left(2-\frac{\pi}{2}\right)\sigma^2$	$2\sigma^2$

表 4.3-2　一些常见离散随机变量的均值，方差和均方值

分 布 簇	PMF $P(k)$	均值 $\mu=E[K]$	方差 σ^2	均方值 $E[K^2]$
伯努利分布	$P_B(k)=\begin{cases}1, & p\\ 0, & q\end{cases}\ q\triangleq 1-p$	p	pq	p
Binomial	$b(k;n,p)=\binom{n}{k}p^kq^{n-k}$	np	npq	$(np)^2+npq$
几何分布①	$\dfrac{1}{1+\mu}\left(\dfrac{\mu}{1+\mu}\right)^k u(k)$	μ	$\mu+\mu^2$	$\mu+2\mu^2$
泊松分布	$\dfrac{\alpha^k}{k!}\mathrm{e}^{-\alpha}u(k)$	α	α	$\alpha^2+\alpha$

联合矩

现在再讨论一下例 4.2-3 中首次提及的那个问题。假设给定两个随机变量 X 和 Y，我们希望测量一下二者的线性关系到底有多“好”。换句话说，也就是通过观测 X 的取值来衡量 Y 的取值。一种极端情况，如果 X 和 Y 是完全独立的，那么 X 的观测值对于 Y 的估计毫无帮助。另一种极端情况，如果 $Y=aX+b$，只要知道了 X 的取值马上就能知道 Y 的取值。但是在实际情况下，两个随机变量既不会完全独立，也不会完全线性相关。基于这个现实非常有必要定义一种测度，可以用来衡量一个随机变量能“告诉”我们多少关于另外一个随机变量的信息，这种测度就叫做联合矩。并不是所有的联合矩都是一样重要的，最重要的一种是二阶联合矩（将很快在后面给出定义）。但是，后面将会看到，其他联合矩在另外一些应用场合也是很重要的，所以在此先讨论一般情况。

定义 4.3-3　X 和 Y 的第 ij 阶联合矩定义如下：

$$\begin{aligned} m_{ij} &\triangleq E[X^iY^j] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^iy^jf_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \end{aligned} \tag{4.3-13}$$

如果 X 和 Y 是离散的，可以利用概率质量函数来计算 μ_{ij}

$$m_{ij} \triangleq \sum_l\sum_m x_l^iy_m^jP_{X,Y}(x_l,y_m) \tag{4.3-14}$$

定义 4.3-4　X 和 Y 的第 ij 阶中心矩定义如下：

$$c_{ij} \triangleq E[(X-\overline{X})^i(Y-\overline{Y})^j] \tag{4.3-15}$$

其中用到了前面的一个定义 $\overline{X}\triangleq E[X]$，对于 $\overline{Y}$ 也类似。矩的阶数为 $i+j$。因此，下面给出的都是二阶矩：

$$\begin{aligned} m_{02} &= E[Y^2] & c_{02} &= E[(Y-\overline{Y})^2] \\ m_{20} &= E[X^2] & c_{20} &= E[(X-\overline{X})^2] \\ m_{11} &= E[XY] & c_{11} &= E[(X-\overline{X})(Y-\overline{Y})] \\ & & &= E[XY]-\overline{X}\,\overline{Y} \\ & & &\triangleq \mathrm{Cov}[X,Y] \end{aligned}$$

① 有的时候几何分布的概率质量函数被写成 $(1-a)a^ku(k)$ 的形式，$0<a<1$，其中 $\mu=a/(1-a)$，$\mu>0$。

m_{11} 和 c_{11} 是两个最重要的联合矩，可以用来衡量预测能力与统计相关性，它们分别被称为 X 和 Y 的协相关与协方差。协相关系数①的定义为

$$\rho \triangleq \frac{c_{11}}{\sqrt{c_{02}c_{20}}} \tag{4.3-16}$$

该定义在4.2节已经介绍过[参见式(4.2-16)]。为了证明协相关系数 $|\rho| \leqslant 1$，考虑如下的非负表达式：

$$E[(\lambda(X-\mu_X)-(Y-\mu_Y))^2] \geqslant 0$$

λ 为任意的实常数。为了验证不等式左边为非负的，我们只需要将其改写为如下的形式：

$$Q(\lambda) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\lambda(x-\mu_X)-(y-\mu_Y)]^2 f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \geqslant 0$$

因此对非负的不等式进行积分结果不可能为负，结论得证。

上一个等式其实是 λ 的一个二次方程。实际上，将该方程展开可得

$$Q(\lambda)=\lambda^2 c_{20}+c_{02}-2\lambda c_{11} \geqslant 0$$

方程 $Q(\lambda)$ 至少有一个实根。因此，它的判别表达式应该满足

$$\left(\frac{c_{11}}{c_{20}}\right)^2-\frac{c_{02}}{c_{20}} \leqslant 0$$

或者

$$c_{11}^2 \leqslant c_{02}c_{20} \tag{4.3-17}$$

因此结论 $|\rho| \leqslant 1$ 得证。

当 $c_{11}^2 = c_{02}c_{20}$ 时 $|\rho| = 1$，也就是

$$E\left[\left(\frac{c_{11}}{c_{20}}(X-\mu_X)-(Y-\mu_Y)\right)^2\right]=0$$

其等价表达式为

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{c_{11}}{c_{20}}(x-\mu_X)-(y-\mu_Y)\right)^2 f_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y=0 \tag{4.3-18}$$

因为 $f_{XY}(x, y)$ 不可能为负，式(4.3-18)的结论表明括号内的表达式处处为0②。因此，通过式(4.3-18)可知，当 $|\rho| = 1$ 时

$$Y=\frac{c_{11}}{c_{20}}(X-\mu_X)+\mu_Y \tag{4.3-19}$$

也就是说 Y 是 X 的一个线性方程，此时 $\mathrm{Cov}[X, Y] = 0$，$\rho = 0$，X 和 Y 不相关。

不相关随机变量的一些特性

(a) 如果 X 和 Y 是不相关的，那么

$$\sigma_{X+Y}^2=\sigma_X^2+\sigma_Y^2 \tag{4.3-20}$$

---

① 注意：用协方差系数或者归一化协方差的称谓可能更合适。

② 在此不考虑那些概率为0的奇异点。为了表达更准确一些，我们应该把“处处”改为“几乎处处”，不过我们经常采用简单的写法。

其中

$$\sigma_{X+Y}^2 \triangleq E[(X+Y)^2]-(E[X+Y])^2$$

(b) 如果 X 和 Y 是独立的，那么它们不相关。

关于(a)的证明，留给读者作为练习；(b)因为 $\mathrm{Cov}[X,Y]=E[XY]-E[X]E[Y]$，我们必须证明 $E[XY]=E[X]E[Y]$，即

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} xf_X(x)\,\mathrm{d}x\int_{-\infty}^{\infty} yf_Y(y)\,\mathrm{d}y \quad (\text{基于独立的假设}) \\ &= E[X]E[Y] \end{aligned}$$

根据独立性假设，结论得证。

例 4.3-4　（线性预测）假设我们希望通过对随机变量 X 的观测来估计随机变量 Y。以图 4.3-2 的数据为例，用一个线性方程来对 Y 进行建模可能更合适

$$Y_P \triangleq \alpha X+\beta \tag{4.3-21}$$

尽管 Y 与 X 之间有联系，但 Y 的取值有可能受其他因素所影响。因此，一般情况下 $|\rho|\neq 1$。很显然 Y 的估计值 Y_P 与真实值之间是有一定误差的。那么我们此时需要做的就是寻找合适的系数 α 和 β，使得下面的均方误差最小：

$$\varepsilon^2 \triangleq E[(Y-Y_P)^2] \tag{4.3-22}$$

这个例题其实就是一个最简单的最优线性估计。从统计学角度，也称为线性回归。

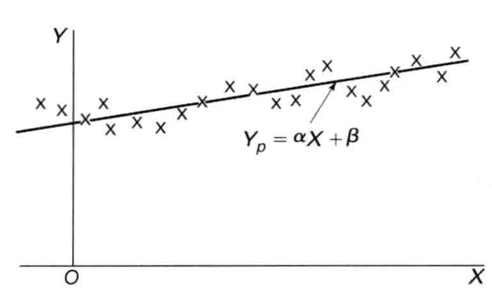


图 4.3-2　数据对(X, Y)的散点图。X 和 Y 之间的关系可以用一条直线来近似

解　将式(4.3-22)展开可得

$$\varepsilon^2 = E[Y^2]-2\alpha\mu_{XY}-2\beta\mu_Y+2\alpha\beta\mu_X+\alpha^2E[X^2]+\beta^2$$

为使 ε 最小，我们对 ε 关于 α 和 β 求偏导

$$\frac{\partial\varepsilon^2}{\partial\alpha}=0,\qquad \frac{\partial\varepsilon^2}{\partial\beta}=0 \tag{4.3-23}$$

用 α_0 和 β_0 来表示最佳系数 α 和 β，利用 α_0 和 β_0 可使 ε 最小化。通过一些代数技巧可得

$$\alpha_0=\frac{\mathrm{Cov}[X,Y]}{\sigma_X^2}=\frac{\rho\sigma_Y}{\sigma_X} \tag{4.3-24a}$$

以及

$$\begin{aligned} \beta_0 &= \overline{Y}-\frac{\mathrm{Cov}[X,Y]}{\sigma_X^2}\overline{X} \\ &= \overline{Y}-\rho\frac{\sigma_Y}{\sigma_X}\overline{X} \end{aligned} \tag{4.3-24b}$$

因此，最优线性估计器为

$$Y_P-\mu_Y=\rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X) \tag{4.3-25}$$

该直线通过了点 (μ_X, μ_Y)。如果将 α_0 和 β_0 代入式(4.3-22)可得最小均方误差 $\varepsilon_{\min}^2$，这也就是习题 4.33。

$$\varepsilon_{\min}^2 = \sigma_Y^2(1-\rho^2) \tag{4.3-26}$$

当 $\rho = 0$ 时还有一些非常有趣的结论。根据式(4.3-25)可知，当 $\rho = 0$ 时，估计值 $Y_P = \mu_Y$，这与 X 毫无关系！这意味着对 X 的观测不会对 Y 的估计带来任何帮助，而且此时最优的估计器就是 $Y_P = \mu_Y$。在例 4.2-3 中也遇到了类似的情况。因此，在线性预测与联合高斯概率密度函数的问题中，用协相关系数来衡量预测能力是比较合适的。在有的时候，两个随机变量不相关貌似二者是独立的。毫无疑问，这种观点来自这么一个结论，即两个随机变量 X 和 Y 独立意味则二者是不相关的。但是，这个结论反过来就不成立了。下面用一个例题来说明。

例 4.3-5 (不相关弱于独立)考虑随机变量 X 和 Y，联合概率质量函数 $P_{X,Y}(x_i, y_j)$ 为

$P_{X,Y}(x_i, y_j)$ 的值

	$x_1=-1$	$x_2=0$	$x_3=+1$
$y_1=0$	0	$\frac{1}{3}$	0
$y_2=1$	$\frac{1}{3}$	0	$\frac{1}{3}$

X 和 Y 不是独立的，因为 $P_{XY}(0, 1) = 0 \neq P_X(0)P_Y(1) = \dfrac{2}{9}$。另外，因为 $\mu_X = 0$，所以 $\text{Cov}(X, Y) = E[XY] - \mu_X\mu_Y = E[XY]$，$m_{11}$ 可计算如下：

$$m_{11} = (-1)(1)\tfrac{1}{3} + (1)(1)\tfrac{1}{3} = 0$$

因此，X 和 Y 不相关，但二者不独立。

$\rho = 0$ 是一个非常重要的特例，此时往往意味着独立。接下来将讨论这种情况。

联合高斯随机变量

如果两个随机变量的概率密度函数满足如下表达式，那么就称其为联合高斯①(联合正态)的

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \exp\left(\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}\right) \tag{4.3-27}$$

这里有 5 个相关参数，分别是 σ_X，σ_Y，μ_X，μ_Y 和 ρ。如果 $\rho = 0$，就有

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

其中

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right) \tag{4.3-28}$$

---

① 联合正态概率密度函数有的时候也称为二维正态概率密度函数，这是为了与多维正态概率密度函数一致。如果不采用矩阵(参见第 5 章)，后者的表达式将非常复杂。

以及

$$f_Y(y)=\frac{1}{\sqrt{2\pi\sigma_Y^2}}\exp\left(-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right) \tag{4.3-29}$$

因此，两个联合高斯随机变量，如果它们不相关（即 $\rho=0$），那它们也是相互独立的。无论 ρ 取什么值，边缘密度函数 $f_X(x)$ 和 $f_Y(y)$ 一般也是正态的。但是反之不成立，即如果 $f_X(x)$ 和 $f_Y(y)$ 是正态的，并不能肯定 X 和 Y 是联合高斯的。

为了加强了解，我们以 CT（计算机 X 射线断层扫描）为例进行说明。这是一种非常流行的 X 射线成像技术，常用于检测人体中的癌细胞以及其他异常细胞。假设存在一个物体，已知它的 X 射线吸收函数为 $f(x,y)\geqslant 0$。这个函数类似于一个联合概率密度函数，即它是一个实函数，非负且很容易被归一化（不过，最后一个性质并不重要）。因此，我们可以在联合概率密度函数 $f_{XY}(x,y)$ 和 X 射线吸收函数 $f(x,y)$ 之间建立一个一一对应的关系。在 CT 检测中，X 射线沿着不同的路径通过一个物体，在一个特定的角度上，所有累计的吸收被测量和记录下来。每个吸收称为一个投影，在特定角度 θ 上所有投影的集合被称为 θ 角度上的轮廓函数。因此，一条直线在 θ 角度上的投影以及与中心位置的偏移量 s 由如下表达式给出［参见图 4.3-3(a)］：

$$f_\theta(s)=\int_{L(s,\theta)} f(x,y)\mathrm{d}l$$

其中，$L(s,\theta)$ 表示一条偏离中心位置的直线上的点，s 表示偏移量，θ 表示偏角，$\mathrm{d}l$ 表示直线 $L(s,\theta)$ 上的微分。如果 s 从最小值变到最大值，那就可以得到这个特定角度上的轮廓函数。通过一种叫做"滤波卷积反投影"的复杂算法，可以利用所有角度上的轮廓函数来得到高质量的人体 X 射线成像。假设我们得到了从 0°～90°的所有轮廓函数，如图 4.3-3(b) 所示，那么可得

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)\mathrm{d}y \qquad \text{（水平面）}$$

$$f_2(y)=\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x \qquad \text{（垂直面）}$$

如果 $f(x,y)$ 是高斯的，那么 $f_1(x)$ 和 $f_2(y)$ 也将是高斯的，这与边缘概率密度函数类似。现在我们探讨一下这种情况：如果把高斯的 $f(x,y)$ 改为非高斯的，在高斯轮廓函数里面能否看到什么变化？如果有变化，就可以给出这么一个结论：边缘函数是高斯的并不意味着联合概率密度函数也是高斯的。在图 4.3-3(c) 中，沿着 45°角从 $a\sim b$ 增加某个物体的吸收率，增加的量为 P，同时也沿着 135°角从 $a'\sim b'$ 减少这个物体的吸收率，减少的量也为 P。由于在水平和垂直方向上，轮廓积分增加和减少的量都是 P，因此 $f_1(x)$ 和 $f_2(y)$ 的净变化为 0，结论得证。假设 P 不是一个过大的值，那么 $f(x,y)$ 沿着 $a'\sim b'$ 的结果为负。做出这个假设的原因在于概率密度函数以及 X 射线的吸收率不可能为负。

为更深刻理解联合高斯分布，我们考虑这么一个非常理想的例子。假设从大量已婚夫妇中随机选择，分别用 X 和 Y 代表丈夫和妻子的身高。假设 X 和 Y 服从高斯分布且相互独立，这只是一种近似，因为人的身高最小不可能为负数，最高要受物理因素制约。在我们的社会中，一般都是高个子选择高个子，矮个子选择矮个子。在这种情况下，X 和 Y 就是正相关的，也即是 $\rho>0$。另一种情况，在某个特定的社会中，可能高个子男性喜欢选择矮个子女性，

而高个子女性却喜欢选择矮个子男性，此时 X 和 Y 就是负相关的，也就是说 $\rho<0$。最后，如果这个社会中男女的婚配完全是随机选择的，那么此时的 ρ 就是 0 或者一个非常小的数①。

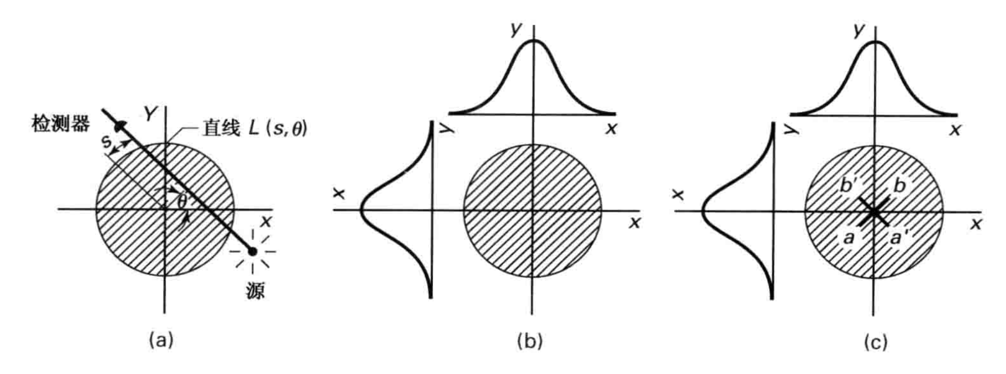


图 4.3-3 通过 CT 成像技术来理解边缘概率密度函数是高斯的，联合概率密度函数不一定是高斯的。(a) 一条直线的投影，偏差是 s，角度为 θ。一个特定角度上所有投影的集合组成了该角度的轮廓函数；(b) 一个联合高斯的X射线物体在水平和垂直方向上都产生了类似高斯分布的轮廓曲线；(c) 在 $a\sim b$ 方向上加上一个恒定的吸收率，在 $a'\sim b'$ 方向上减去一个同样的吸收率，轮廓函数不变，但实际上的吸收率已不再是高斯的

*** 联合高斯概率密度函数的等密度线②** 如果把 $f_{XY}(x,y)$ 设定为常数，那么需要找出 xy 平面上的焦点。很显然，指数部分如果为常数(假设为 a^2)，那么 $f_{XY}(x,y)$ 也就是常数

$$\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2=a^2$$

上面这个等式是一个椭圆方程，焦点分别是 $x=\mu_X$ 和 $y=\mu_Y$。为简单起见，设 $\mu_X=\mu_Y=0$。当 $\rho=0$ 时，椭圆的长轴和短轴分别平行于 x 轴和 y 轴，我们知道此时 X 和 Y 是独立的。如果 $\rho=0$ 且 $\sigma_X=\sigma_Y$，此时椭圆退化为一个圆。图 4.3-4 给出了许多示例。

从式(4.3-27)得到的边缘概率密度函数 $f_X(x)$ 和 $f_Y(y)$ 不依赖于参数 ρ，这有些让人惊奇。为了证明这个结论，我们计算如下：

$$f_X(x)=\int_{-\infty}^{\infty}f_{XY}(x,y)\mathrm{d}y$$

为简单起见，$\mu_X=\mu_Y=0$。上面这个积分看上去有些复杂，但按照下面三个步骤计算起来就相对容易：

1. 把所有与 y 不相关的因子从积分表达式中提取出来。
2. 把指数部分凑成完全平方表达式(参见附录 A 中“完全平方表达式”)。
3. 对于 $b>0$ 和实数 y，有如下的结论成立：

$$\frac{1}{\sqrt{2\pi b^2}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(\frac{y-a}{b}\right)^2\right]\mathrm{d}y=1$$

实际上，经过步骤 2 就可以得到

① 要想观测到相关性为 0 的两个随机变量，这在实际的统计学中是非常不可能的，即使二者在理论上是不相关的。有些夸夸其谈的人试图用微小的随机相关性来证明一个事件相关性为 0，这种思路其实并不合理。

② 第一次学习时可以不用阅读带 * 号的内容。

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_X}\exp\left[-\frac{1}{2}\left(\frac{x}{\sigma_X}\right)^2\right]\times\left\{\frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}}\int_{-\infty}^{\infty}\exp\left[\frac{-(y-\rho x\sigma_Y/\sigma_X)^2}{2\sigma_Y^2(1-\rho^2)}\right]\mathrm{d}y\right\} \tag{4.3-30}$$

上式大括号中的表达式值为 1，因此有

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_X}\exp\left[-\frac{1}{2}\left(\frac{x}{\sigma_X}\right)^2\right] \tag{4.3-31}$$

通过同样的过程可以计算出 $f_Y(y)$

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left[-\frac{1}{2}\left(\frac{y}{\sigma_Y}\right)^2\right] \tag{4.3-32}$$

如我们前面所说的，如果 $\rho=0$，那么 X 和 Y 是相互独立的。另一方面，如果 $\rho\to\pm1$，X 和 Y 趋向于线性相关。为简单起见，令 $\sigma_X=\sigma_Y\triangleq\sigma$，$\mu_X=\mu_Y=0$，那么等密度线变为

$$x^2-2\rho xy+y^2=c^2\sigma^2$$

当 $\rho>0$ 时这是一个与 x 轴成 45°角的椭圆，当 $\rho<0$ 时成 135°角。通过下面的坐标变换，可以得到一个在 x-y 平面上旋转 45°角的坐标系。

$$v=\frac{x+y}{\sqrt{2}}\quad w=\frac{x-y}{\sqrt{2}}$$

此时等密度线变为

$$v^2[1-\rho]+w^2[1+\rho]=\sigma^2c^2$$

这个椭圆的长轴和短轴分别与 v 轴和 w 轴平行。如果 $\rho>0$，长轴与 v 轴平行；如果 $\rho<0$，长轴与 w 轴平行。当 $\rho\to\pm1$ 时，长轴变得无限长，此时概率密度函数集中收拢为直线 $y=x(\rho\to1)$ 或 $y=-x(\rho\to-1)$。

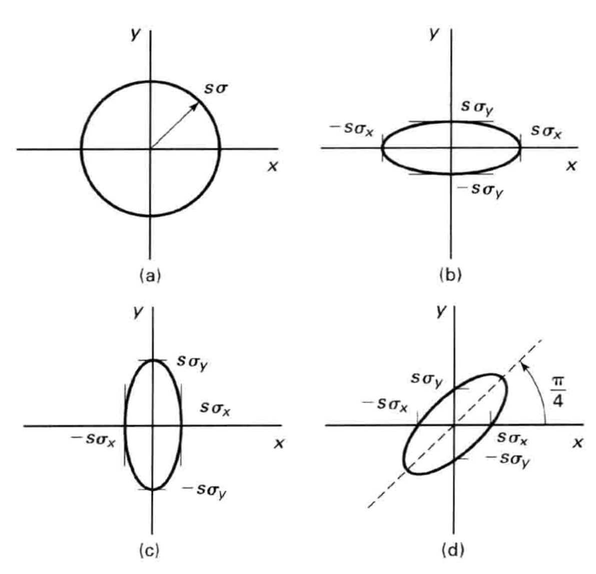


图 4.3-4　联合高斯密度函数的等密度线（$\bar{X}=\bar{Y}=0$）。(a) $\sigma_X=\sigma_Y$，$\rho=0$；(b) $\sigma_X>\sigma_Y$，$\rho=0$；(c) $\sigma_X<\sigma_Y$，$\rho=0$；(d) $\sigma_X=\sigma_Y$，$\rho>0$

最后，引入下面两个新的随机变量

$$V \triangleq (X+Y)/\sqrt{2}$$

$$W \triangleq (X-Y)/\sqrt{2}$$

当$\rho \to 1$时有

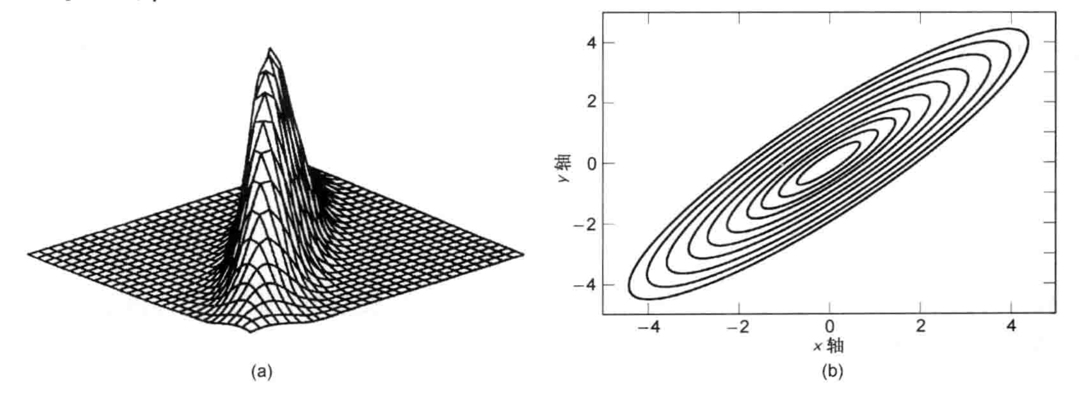
$$f_{XY}(x,y) \to \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right] \times \delta(y-x)$$

或者，等价的表达式为

$$f_{XY}(x,y) \to \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2\right] \times \delta(y-x)$$

联合高斯的概率密度函数退化为单变量的函数(即直线$y = x$)，这是因为当$\rho \to 1$时X等同于Y。我们将详细推导过程留给学生作为练习。

图4.3-5给出了计算机生成的联合高斯概率密度函数和它的等密度线，其中$\mu_X = \mu_Y = 0$，$\sigma_X = \sigma_Y = 2$，$\rho = 0.9$。



4.3-5 (a) $\overline{X} = \overline{Y} = 0$，$\sigma_X = \sigma_Y = 2$，$\rho = 0.9$时的联合高斯概率密度函数；(b)等密度线示意图

4.4 切比雪夫和施瓦茨不等式

切比雪夫①不等式给出了一个界限，告诉我们一个随机变量X偏离它的均值μ_X的概率。

定理4.4-1 (切比雪夫不等式)设X为任意一个随机变量，均值为μ_X，方差为有限值σ^2。那么对于任意的$\delta > 0$

$$P[|X-\mu_X| \geqslant \delta] \leqslant \frac{\sigma^2}{\delta^2} \tag{4.4-1}$$

证明 通过下面的步骤可以很容易的证明式(4.4-1)

$$\sigma^2 \triangleq \int_{-\infty}^{\infty}(x-\overline{X})^2 f_X(x)\,\mathrm{d}x \geqslant \int_{|x-\overline{X}|\geqslant\delta}(x-\overline{X})^2 f_X(x)\,\mathrm{d}x$$

$$\geqslant \delta^2\int_{|x-\overline{X}|\geqslant\delta} f_X(x)\,\mathrm{d}x$$

$$= \delta^2 P[|X-\overline{X}| \geqslant \delta]$$

① Pafnuty L. Chebyshev(1821 – 1894)，即切比雪夫，俄罗斯数学家。

因为 $\{|X-\overline{X}|\geqslant\delta\}\cup\{|X-\overline{X}|<\delta\}=\Omega$($\Omega$表示一个确定事件)，而且这两个取并集的事件是不相交的，因此有

$$P[|X-\overline{X}|<\delta]\geqslant 1-\frac{\sigma^2}{\delta^2} \tag{4.4-2}$$

有的时候，用 σ 来表示 δ 更方便一些，即 $\delta\triangleq k\sigma$，其中 k 是一个常数①。此时式(4.4-1)和式(4.4-2)分别成为

$$P[|X-\overline{X}|\geqslant k\sigma]\leqslant\frac{1}{k^2} \tag{4.4-3}$$

$$P[|X-\overline{X}|<k\sigma]\geqslant 1-\frac{1}{k^2} \tag{4.4-4}$$

例4.4-1　(一个高斯随机变量与均值的偏差)设 X: $N(\mu_X,\sigma^2)$。请把 $P[|X-\mu_X|<k\sigma]$ 和 $P[|X-\mu_X|\geqslant k\sigma]$ 同切比雪夫限(CB)相比较。

解　根据式(2.4-14d)和式(2.4-14e)，$P[|X-\mu_X|<k\sigma]=2\text{erf}(k)$，$P[|X-\mu_X|\geqslant k\sigma]=1-2\text{erf}(k)$，其中函数 erf($k$) 的定义参见式(2.4-12)。再通过表2.4-1，式(4.4-3)和式(4.4-4)，结果如表4.4-1所示。

从表4.4-1可以看出，切比雪夫限并不是很好，但是对于任意随机变量 X，只要 σ^2 存在，就需要记住有这个界限存在。

还存在着切比雪夫不等式的一些扩展②，我们接下来将学习其中一个。

表4.4-1

k	$P[\lvert X-\overline{X}\rvert<k\sigma]$	CB	$P[\lvert X-\overline{X}\rvert>k\sigma]$	CB
0	0 PMF $P(k)$	0	1	1
0.5	0.383	0	0.617	1
1.0	0.683	0	0.317	1
1.5	0.866	0.556	0.134	0.444
2.0	0.955	0.750	0.045	0.250
2.5	0.988	0.840	0.012	0.160
3.0	0.997	0.889	0.003	0.111

马尔可夫不等式

考虑一个非负的随机变量 X，其中 $x<0$ 时 $f_X(x)=0$，此时可以运用马尔可夫不等式

$$P[X\geqslant\delta]\leqslant\frac{E[X]}{\delta} \tag{4.4-5}$$

在切比雪夫限中同时包含了 X 的均值和方差，而马尔可夫限中只涉及了 X 的均值。

式(4.4-5)的证明

$$E[X]=\int_0^{\infty}xf_X(x)\,\mathrm{d}x\geqslant\int_{\delta}^{\infty}xf_X(x)\,\mathrm{d}x\geqslant\delta\int_{\delta}^{\infty}f_X(x)\,\mathrm{d}x$$
$$\geqslant\delta P[X\geqslant\delta]$$

稍做变换即可得到式(4.4-5)。式(4.4-5)给出了一个界限，告诉我们总体中大于 δ 的那部分的概率。

---

① 当 k 或 δ 很小时，切比雪夫不等式就不太有用。

② 参见 Davenport[4-2, p.256]。

例 4.4-2 (劣质电阻器)假设生产出了一批低质量的 1000 Ω 电阻，通过测试与统计分析这批电阻的均值确实为 1000 Ω，但是有许多电阻偏离这个均值。假设电阻超过 1500 Ω 为废品，请计算这批电阻中废品的比例最大为多少?

解 因为 $\mu_X = 1000$，$\delta = 1500$，我们可得

$$P[X \geqslant 1500] \leqslant \frac{1000}{1500} = 0.67$$

因此，排除其他因素，生产厂家敢确保废品率不会超过 67% 。

施瓦茨不等式

在例 4.3-17 中已经出现过施瓦茨[①]不等式的概率表达式，在这里我们再次给出表达式如下：

$$\mathrm{Cov}^2[X,Y] \leqslant E[(X-\mu_X)^2]E[(Y-\mu_Y)^2]$$

当且仅当 Y 是 X 的线性函数时，上式才能取等号。对不等式两边同时求平方根可知，两个随机变量方差乘积的平方根是它们协方差幅度的上限，即

$$|\mathrm{Cov}[X,Y]| \leqslant \left(\sigma_X^2\sigma_Y^2\right)^{1/2}$$

接下来，将要了解施瓦茨表达式的另外一种表达式，这在信号处理和随机过程中经常用到。考虑两个非随机(即确定性)函数 $\boldsymbol{h}$ 和 $\boldsymbol{g}$，$\boldsymbol{h}$ 和 $\boldsymbol{g}$ 不一定是实函数。当积分存在时，定义函数 f 的范数为

$$\|f\| \triangleq \left(\int_{-\infty}^{\infty} |f(x)|^2\,\mathrm{d}x\right)^{1/2} \tag{4.4-6}$$

同样地，用 $(\boldsymbol{h},\boldsymbol{g})$ 表示 $\boldsymbol{h}$ 和 $\boldsymbol{g}$ 的点积或内积，定义如下：

$$\begin{aligned}(h,g) &\triangleq \int_{-\infty}^{\infty} h(x)g^*(x)\,\mathrm{d}x \\ &= (g,h)^*\end{aligned} \tag{4.4-7}$$

施瓦茨表达式的确定性表达式此时为

$$|(h,g)| \leqslant \|h\|\|g\| \tag{4.4-8}$$

当且仅当 $\boldsymbol{h}$ 正比于 $\boldsymbol{g}$ 时，即 $\boldsymbol{h}(x) = \boldsymbol{a}\boldsymbol{g}(x)$，$\boldsymbol{a}$ 为一个常数，上述不等式取等号。为了证明式(4.4-8)，可将 $\lambda \boldsymbol{h}(x) + \boldsymbol{g}(x)$ 看成变量 λ 的函数

$$\|\lambda h(x) + g(x)\|^2 = |\lambda|^2\|h\|^2 + \lambda(h,g) + \lambda^*(h,g)^* + \|g\|^2 \geqslant 0 \tag{4.4-9}$$

如果令

$$\lambda = -\frac{(h,g)^*}{\|h\|^2} \tag{4.4-10}$$

那么式(4.4-8)得证。对于实随机变量 $\boldsymbol{X}$，$\boldsymbol{h}(\boldsymbol{X})$ 和 $\boldsymbol{g}(\boldsymbol{X})$ 是 $\boldsymbol{X}$ 的实函数，此时式(4.4-8)仍然成立，此时范数和内积的定义需要修改为如下形式：

$$\|h\|^2 \triangleq \int_{-\infty}^{\infty} h^2(x)f_X(x)\,\mathrm{d}x = E[h^2(X)] \tag{4.4-11}$$

---

① H. Amandus Schwarz(1843 – 1921)，即施瓦茨，德国数学家。

$$(h,g) \triangleq \int_{-\infty}^{\infty} h(x)g(x)f_X(x)\,\mathrm{d}x = E[h(X)g(X)] \tag{4.4-12}$$

此时可得

$$|E[h(X)g(X)]| \leqslant (E[h^2(X)])^{1/2}(E[g^2(X)])^{1/2} \tag{4.4-13}$$

大数定理　切比雪夫不等式的一个重要应用是用来证明弱大数定理(LLN)，即告诉我们在什么条件下样本均值收敛于总体均值。

例 4.4-3　(弱大数定理)设 $X_1, \cdots, X_n$ 为独立同分布的随机变量，均值为 μ_X，方差为 σ_X^2。假设我们并不知道 μ_X（或 σ_X）的具体数值，考虑如下的样本均值估计器①：

$$\hat{\mu}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i$$

用上面的结果作为均值 μ_X 的估计值。利用切比雪夫表达式，可以证明 $\hat{\mu}_n$ 是 μ_X 的渐进完美估计值。首先，可以计算 $\hat{\mu}_n$ 的数学期望如下：

$$\begin{aligned} E[\hat{\mu}_n] &= \frac{1}{n}\sum_{i=1}^{n} E[X_i] \\ &= \frac{1}{n} n\mu_X \\ &= \mu_X \end{aligned}$$

紧接着，计算 $\hat{\mu}_n$ 的方差如下：

$$\begin{aligned} \mathrm{Var}[\hat{\mu}_n] &= \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] \\ &= \left(\frac{1}{n^2}\right) n\sigma_X^2 \\ &= \frac{1}{n}\sigma_X^2 \end{aligned}$$

根据切比雪夫不等式[参见式(4.4-1)]可得

$$P[|\hat{\mu}_n - \mu_X| \geqslant \delta|] \leqslant \sigma_X^2 / n\delta^2$$

很显然，对于任意的 $\delta > 0$，可以取足够大的 n 来使得表达式成立。因此对于任意 $\delta > 0$，有下面的结论：

$$\lim_{n\to\infty} P[|\hat{\mu}_n - \mu_X| \geqslant \delta] = 0$$

在这种情况下，如果 δ 非常小，那么 n 就需要取很大的值才能保证事件 $\{|\hat{\mu}_n - \mu_X| \geqslant \delta\}$ 的概率足够小。这种类型的收敛称为依概率收敛，在第 8 章我们还会详细讲解。

大数定理是估计 μ_X 的一个理论依据。根据大数定理，一个实验人员就可以用 n 次测量的样本均值来作为未知数学期望 $E[X] = \mu_X$ 的一种估计。

有的时候，也可以由概率密度函数来推导出不等式。下面用一个例子来进行说明，这个例题是由 Yongyi Yang 设计的。

---

① 估计器是观测值 $X_1, X_2, \cdots, X_n$ 的一个函数，可用来估计分布的某种参量。估计器也是一个随机变量。如果把一个特定的值(即一种实现)输入估计器，那么得到的输出结果就称为一个估计值。在第 6 章将会详细讨论有关估计器的内容。

例 4.4-4 (对称随机变量)设实随机变量 X 的概率密度函数满足 $f_X(x)=f_X(-x)$，因此 X 关于零点对称分布。试证明当 $\mathrm{Var}(|X|)=0$ 时，不等式 $\sigma_X \geqslant E[|X|]$ 取等号。

解 设 $Y \triangleq |X|$。因为 $\mu_X=0$，因此 $E[Y^2]=E[X^2]=\mu_X^2+\sigma_X^2=\sigma_X^2$。同样地，$E[Y^2]=\mu_Y^2+\sigma_Y^2=E^2[|X|]+\sigma_Y^2=\sigma_X^2$。因为 $\sigma_Y^2 \geqslant 0$，只有当 $\sigma_Y^2=0$ 时，$E^2[|X|] \leqslant \sigma_X^2$ 才能取等，结论得证。进一步地，如果 X 的概率密度函数为 $f_X(x)=\dfrac{1}{2}[\delta(x-a)+\delta(x+a)]$，$a$ 是一个正常数，此时 $Y=a$，$\sigma_Y=0$，$E[|X|]=\sigma_X$。

还有另外一种不等式称为切尔诺夫限(Chernoff Bound)。在 4.6 节我们将会讨论这种界限，不过接下来先学习矩量生成函数(moment - generating function) $M(t)$。

4.5 矩量生成函数

一个随机变量 X 的矩量生成函数(MGF)如果存在，其定义如下①：

$$M(t) \triangleq E[\mathrm{e}^{tX}] \tag{4.5-1}$$

$$= \int_{-\infty}^{\infty} \mathrm{e}^{tx} f_X(x)\,\mathrm{d}x \tag{4.5-2}$$

其中 t 是一个复变量。

对于离散随机变量，我们可用概率质量函数来定义 $M(t)$

$$M(t)=\sum_i \mathrm{e}^{tx_i} P_X(x_i) \tag{4.5-3}$$

从式(4.5-2)可以看出，除了指数部分少了一个负号，矩量生成函数就是概率密度函数的双边拉普拉斯变换。而且拉普拉斯变换还存在着一个逆变换的公式，因此如果已知 $M(t)$，等同于 $f_X(x)$ 已知，反之亦然。

在这里介绍 $M(t)$ 的原因在于：(1)可以用来很方便的计算 X 的矩；(2)可以根据实验测量中获得的矩来估计 $f_X(x)$；(3)可以用来解决其他很多问题，比如计算随机变量的和；(4)这是一个非常重要的解析工具，可以用来描述许多基本结论，比如中心极限定理②。

如果将 e^{tX} 展开并取数学期望，可得

$$\begin{aligned} E[\mathrm{e}^{tX}] &= E\left[1+tX+\frac{(tX)^2}{2!}+\cdots+\frac{(tX)^n}{n!}+\cdots\right] \\ &= 1+tm_1+\frac{t^2}{2!}m_2+\cdots+\frac{t^n}{n!}m_n+\cdots \end{aligned} \tag{4.5-4}$$

矩 m_i 有可能是不存在的，比如，对于柯西概率密度函数，除了一阶矩之外的矩都不存在，此时 $M(t)$ 就有可能不存在。另一方面，如果 $M(t)$ 存在，那么通过微分就可以很容易的计算任意的矩。实际上，如果采用下面的记号：

$$M^{(k)}(0) \triangleq \left.\frac{d^k}{\mathrm{d}t^k}(M(t))\right|_{t=0}$$

① 专用术语会有稍许不同(参见 Feller[4-1]，p.411)。
② 在 4.7 节将会讨论中心极限定理。

那么有

$$m_k = M^{(k)}(0), \qquad k = 0, 1, \cdots \tag{4.5-5}$$

例 4.5-1　（高斯分布的矩量生成函数）设 X：$N(\mu, \sigma^2)$，它的矩量生成函数由如下表达式给出：

$$M_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \mathrm{e}^{tx}\, \mathrm{d}x \tag{4.5-6}$$

对指数部分采用"完全平方法"①，我们可以把式(4.5-6)写为

$$\begin{aligned} M_X(t) &= \exp[\mu t + \sigma^2 t^2/2] \\ &\times \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x - (\mu + \sigma^2 t))^2\right) \mathrm{d}x \end{aligned}$$

其中，上式第 2 行的因子为 1，这是一个高斯概率密度函数的积分，其均值为 $\mu + \sigma^2 t$，方差为 σ^2。因此高斯矩量生成函数为

$$M_X(t) = \exp(\mu t + \sigma^2 t^2/2) \tag{4.5-7}$$

从这个结果可以知道

$$M_X^{(1)}(0) = \mu$$
$$M_X^{(2)}(0) = \mu^2 + \sigma^2$$

例 4.5-2　（二项分布的矩量生成函数）设 B 是一个参量为 n 和 p 的二项分布随机变量，n 是实验的次数，p 是每次实验成功的概率，$q = 1 - p$。那么矩量生成函数为

$$\begin{aligned} M_B(t) &= \sum_{k=0}^{n} \mathrm{e}^{tk} \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} [\mathrm{e}^t p]^k q^{n-k} \\ &= (p\mathrm{e}^t + q)^n \end{aligned} \tag{4.5-8}$$

由上述结果可得

$$\begin{aligned} M_B^{(1)}(0) &= np = \mu \\ M_B^{(2)}(0) &= \{np\mathrm{e}^t(p\mathrm{e}^t + q)^{n-1} + n(n-1)p^2\mathrm{e}^{2t}(p\mathrm{e}^t + q)^{n-2}\}_{t=0} \\ &= npq + \mu^2 \end{aligned} \tag{4.5-9}$$

因此有

$$\mathrm{Var}[B] = npq \tag{4.5-10}$$

例 4.5-3　（几何分布的矩量生成函数）设 X 服从几何分布，其概率质量函数为 $P_X(n) = a^n(1-a)u(n)$，$n = 0, 1, 2, \cdots, 0 < a < 1$。因此矩量生成函数可以计算为

$$M_X(t) = \sum_{n=0}^{\infty} (1-a)a^n \mathrm{e}^{tn}$$

---

① 参见附录 A 中"完全平方法"。

$$= (1-a)\sum_{n=0}^{\infty}(ae^t)^n = \frac{1-a}{1-ae^t}$$

因此，均值 $\mu = M'_X(0) = (1-a)(1-ae^t)^{-2}ae^{-t}|_{t=0} = a/(1-a)$。

我们看到，如果所有的矩都存在且已知的话，那么 $M(t)$ 也是已知的[参见式(4.5-4)和式(4.5-2)]。因为拉普拉斯变换可以将 $M_X(t)$ 和 $f_X(x)$ 联系起来，如果二者存在的话①，由 $M_X(t)$ 来求 $f_X(x)$ 起码在理论上是可行的。在实际中，如果需要求随机变量 X 的概率密度函数，X_i 表示对 X 的第 i 次观测，那么可以用如下的方法来估计 X 的第 r 阶矩

$$\widehat{m}_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r \tag{4.5-11}$$

其中 $\hat{m}_r$ 是一个随机变量，表示 r 阶矩估计器，n 表示观测的次数。尽管 $\hat{m}_r$ 是一个随机变量，随着 n 的增大其方差会变得越来越小。当 n 足够大的时候，就可以认为 $\hat{m}_r$ 同 m_r 足够接近(m_r 是一个确定数，不是一个随机变量)。

两个随机变量 X 和 Y 的矩量生成函数定义如下：

$$\begin{aligned} M_{XY}(t_1,t_2) &\triangleq E[e^{(t_1X+t_2Y)}] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(t_1x+t_2y)f_{XY}(x,y)\,dx\,dy \end{aligned} \tag{4.5-12}$$

同式(4.5-4)的步骤类似，通过幂级数展开可得

$$M_{XY}(t_1,t_2) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{t_1^it_2^j}{i!j!}m_{ij} \tag{4.5-13}$$

其中 m_{ij} 的定义如式(4.3-13)所示。

$$M_{XY}^{(l,n)}(0,0) \triangleq \left.\frac{\partial^{l+n}M_{XY}(t_1,t_2)}{\partial t_1^l\partial t_2^n}\right|_{t_1=t_2=0}$$

采用上面的表达符号，从式(4.5-12)或式(4.5-13)可得

$$m_{ln} = M_{XY}^{(l,n)}(0,0) \tag{4.5-14}$$

另外，还有下面一些有用的结论

$$M_{XY}^{(1,0)}(0,0) = \mu_X, \qquad M_{XY}^{(0,1)}(0,0) = \mu_Y \tag{4.5-15}$$

$$M_{XY}^{(2,0)}(0,0) = E[X^2], \qquad M_{XY}^{(0,2)}(0,0) = E[Y^2] \tag{4.5-16}$$

$$M_{XY}^{(1,1)}(0,0) = m_{11} = \text{Cov}[X,Y] + \mu_X\,\mu_Y \tag{4.5-17}$$

4.6 切尔诺夫限

切尔诺夫限(Chernoff bound)给出了尾概率 $P[X \geqslant a]$ 的一个上限，其中 a 是一个事先给定的常数。对于任意的 $t>0$，$u(x-a) \leqslant e^{t(x-a)}$ 始终成立。假设 X 是一个连续随机变量，那么有

① 对于有些分布并不是所有的矩都存在。比如，前面提及的柯西分布，除了一阶矩以外都是不存在的。

$$\begin{aligned}P[X \geqslant a] &= \int_a^{\infty} f_X(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f_X(x)u(x-a)\,\mathrm{d}x\end{aligned} \tag{4.6-1}$$

对上式观测可知，当 $t > 0$ 时下面的不等式成立

$$P[X \geqslant a] \leqslant \int_{-\infty}^{\infty} f_X(x)\mathrm{e}^{t(x-a)}\,\mathrm{d}x \tag{4.6-2}$$

但是，通过式(4.5-2)可知

$$\int_{-\infty}^{\infty} f_X(x)\mathrm{e}^{t(x-a)}\,\mathrm{d}x = \mathrm{e}^{-at}M_X(t) \tag{4.6-3}$$

在矩量生成函数中加入下标是为了强调它与 X 的联系。通过式(4.6-3)和式(4.6-2)可得

$$P[X \geqslant a] \leqslant \mathrm{e}^{-at}M_X(t) \tag{4.6-4}$$

求出上式右边关于 t 的最小值，这个最小值就是这个不等式的最紧约束，也称为切尔诺夫限。接下来，我们将通过一些例题来讲解。

例 4.6-1　(高斯分布的切尔诺夫限)设 X: $N(\mu, \sigma^2)$，求 $P[X \geqslant a]$ 的切尔诺夫限，其中 $a > \mu$。利用式(4.5-7)和式(4.6-3)可得

$$P[X \geqslant a] \leqslant \mathrm{e}^{-(a-\mu)t+\sigma^2 t^2/2}$$

对不等式右边关于 t 求导求最小值，此时 $t = (a-\mu)/\sigma^2$。因此切尔诺夫限为

$$P[X \geqslant a] \leqslant \mathrm{e}^{-(a-\mu)^2/2\sigma^2} \tag{4.6-5}$$

离散随机变量同样也存在切尔诺夫限。例如，假设随机变量 X 的取值为 $X = i$, $i = 0, 1, 2, \cdots$，概率值为 $P[X = i] \triangleq P_X(i)$。对于任意的整数 n 和 k，定义如下：

$$u(n-k) = \begin{cases} 1, & n \geqslant k \\ 0, & \text{其他} \end{cases}$$

根据上面这个定义可得

$$\begin{aligned}P[X \geqslant k] &= \sum_{n=k}^{\infty} P_X(n) \\ &= \sum_{n=0}^{\infty} P_X(n)u(n-k) \\ &\leqslant \sum_{n=0}^{\infty} P_X(n)\mathrm{e}^{t(n-k)} \qquad t \geqslant 0\end{aligned}$$

其中上面最后一步的推导用到了下面的结论：

$$\mathrm{e}^{t(n-k)} \geqslant u(n-k), \qquad t \geqslant 0$$

我们注意到

$$\begin{aligned}\sum_{n=0}^{\infty} P_X(n)\mathrm{e}^{t(n-k)} &= \mathrm{e}^{-tk}\sum_{n=0}^{\infty} P_X(n)\mathrm{e}^{tn} \\ &= \mathrm{e}^{-tk}M_X(t) \qquad \text{[(根据式 4.5-3)]}\end{aligned}$$

因此可得下面的结论

$$P[X \geqslant k] \leqslant \mathrm{e}^{-tk} M_X(t) \tag{4.6-6}$$

同前面类似，对式(4.6-6)右边取最小值就可以计算出切尔诺夫限。我们仍然用一个例题来讲解。

例 4.6-2 （泊松分布的切尔诺夫限）设 X 是服从泊松分布的随机变量，分布参数 $a > 0$。试计算 $P_X(k)$ 的切尔诺夫限，其中 $k > a$。根据习题 4.39 可知矩量生成函数为

$$M_X(t) = \mathrm{e}^{a[\mathrm{e}^t - 1]}$$

以及

$$\mathrm{e}^{-tk} M_X(t) = \mathrm{e}^{-a} \mathrm{e}^{[a\mathrm{e}^t - kt]}$$

令导数为 0

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{e}^{-tk} M_X(t)] = 0$$

令 $t = t_m$ 时取最小值，那么有

$$t_m = \ln \frac{k}{a}$$

当 $a = 2$, $k = 3$ 时，有下面结论成立

$$\begin{aligned} P[X \geqslant 5] &\leqslant \mathrm{e}^{-2} \exp[5 - 5\ln(5/2)] \\ &\leqslant 0.2 \end{aligned}$$

4.7 特征函数

如果我们把式(4.5-1)中的参数 t 换成 $\mathrm{j}\omega$，其中 $\mathrm{j} \triangleq \sqrt{-1}$，那么就可以得到 X 的特征函数(CF)，具体定义如下：

$$\begin{aligned} \Phi_X(\omega) &\triangleq E[\mathrm{e}^{\mathrm{j}\omega X}] \\ &= \int_{-\infty}^{\infty} f_X(x) \mathrm{e}^{\mathrm{j}\omega x}\, \mathrm{d}x \end{aligned} \tag{4.7-1}$$

上面这个公式除了在指数部分有个负号外，其实这就是 $f_X(x)$ 的傅里叶变换。对于离散随机变量，可以利用它的概率质量函数来定义 $\Phi_X(\omega)$

$$\Phi_X(\omega) = \sum_i \mathrm{e}^{\mathrm{j}\omega x_i} P_X(x_i) \tag{4.7-2}$$

特征函数具有矩量生成函数的所有特性。傅里叶变换在统计通信理论中应用非常广泛，而且由于式(4.7-1)的逆变换一般都比较容易得到(可以通过直接积分法或者查询傅里叶变换表，参见参考文献[4-7])，因此在解决独立随机变量之和的问题中常常会用到特征函数。我们已经知道独立随机变量之和的概率密度函数是它们各自概率密度函数的卷积。因此如果有 $Z = X_1 + \cdots + X_N$，其中 X_i 为独立的随机变量，$i = 1, \cdots, N$，那么 Z 的概率密度函数为

$$f_Z(z) = f_{X_1}(z) * f_{X_2}(z) * \cdots * f_{X_N}(z) \tag{4.7-3}$$

可以看出，得到的结果仍然是卷积形式。

实际上，计算式(4.7-3)是比较麻烦的。但是，我们通过对傅里叶变换的学习可知，卷积的傅里叶变换等于各个独立变换的乘积。接下来用一些例子来进行说明。

例 4.7-1　(和的特征函数)设 $Z \triangleq X_1 + X_2$，各自的概率密度函数分别为 $f_{X_1}(x)$，$f_{X_2}(x)$ 和 $f_Z(x)$。试证明 $\Phi_Z(\omega) = \Phi_{X_1}(\omega)\Phi_{X_2}(\omega)$。

解　根据 3.3 节的结论[参见式(3.3-15)]可知

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z-x)\,\mathrm{d}x$$

相应的特征函数为

$$\begin{aligned}\Phi_Z(\omega) &= \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{j}\omega z}\left[\int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z-x)\,\mathrm{d}x\right]\mathrm{d}z \\ &= \int_{-\infty}^{\infty} f_{X_1}(x)\int_{-\infty}^{\infty} f_{X_2}(z-x)\mathrm{e}^{\mathrm{j}\omega z}\mathrm{d}x\,\mathrm{d}z\end{aligned}$$

进行变量替换 $\alpha \triangleq z - x$，我们可以得到变量之和 Z 的特征函数

$$\Phi_Z(\omega) = \Phi_{X_1}(\omega)\Phi_{X_2}(\omega)$$

利用数学归纳法，可以将这个结论可以扩展到 N 个随机变量的情况。如果 $Z = X_1 + \cdots + X_N$，那么 Z 的特征函数为

$$\Phi_Z = \Phi_{X_1}(\omega)\Phi_{X_2}(\omega)\cdots\Phi_{X_N}(\omega)$$

例 4.7-2　(独立同分布随机变量之和的特征函数)设 X_i 是一个独立同分布的随机序列，且 X: $N(0, 1)$，$i = 1, \cdots, N$。试计算下面表达式的概率密度函数：

$$Z \triangleq \sum_{i=1}^{N} X_i$$

解　可以用式(4.7-3)来计算 Z 的概率密度函数。另一方面，用 $\Phi_{X_i}(\omega)$ 来表示 X_i 的特征函数，那么有如下结论：

$$\Phi_Z(\omega) = \Phi_{X_1}(\omega) \times \cdots \times \Phi_{X_N}(\omega) \tag{4.7-4}$$

但是，因为 X_i 是服从 $N(0, 1)$ 的独立同分布随机变量，因此所有 X_i 的特征函数都是相同的，因此我们定义 $\Phi_X(\omega) \triangleq \Phi_{X_1}(\omega) = \cdots = \Phi_{X_N}(\omega)$。所以有

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}\mathrm{e}^{\mathrm{j}\omega x}\,\mathrm{d}x \tag{4.7-5}$$

通过对指数部分凑完全平方表达式，可以得到

$$\begin{aligned}\Phi_X(\omega) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}[x^2 - 2\mathrm{j}\omega x + (\mathrm{j}\omega)^2 - (\mathrm{j}\omega)^2]}\,\mathrm{d}x \\ &= \mathrm{e}^{-\frac{\omega^2}{2}}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(x-\mathrm{j}\omega)^2}\,\mathrm{d}x\end{aligned}$$

可以把上面的积分值看成“高斯概率密度函数”下方的面积，而且该分布的“均值”为 $\mathrm{j}\omega$，因此积分结果为 1①。因此可得 X 的特征函数为

$$\Phi_X(\omega) = \mathrm{e}^{-\frac{\omega^2}{2}}$$

所以 Z 的特征函数为

$$\Phi_Z(\omega) = [\Phi_X(\omega)]^n = \mathrm{e}^{-\frac{1}{2}n\omega^2} \tag{4.7-6}$$

---

① 如果对这个结果不很理解，可以在复平面上对这个积分进行严格的描述。

通过 $\Phi_Z(\omega)$ 的形式可知 $f_Z(z)$ 一定也是高斯的。为了计算 $f_Z(z)$，我们将用到傅里叶逆变换的公式

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_Z(\omega)\mathrm{e}^{-\mathrm{j}\omega z}\mathrm{d}\omega \tag{4.7-7}$$

乘以适当的因子后，将式(4.7-6)代入式(4.7-7)可得

$$f_Z(z) = \frac{1}{\sqrt{2\pi n}}\mathrm{e}^{-\frac{1}{2}(z^2/n)}$$

因此可以看出 $f_Z(z)$ 确实是高斯的。Z 的方差是 n，均值为 0。

例 4.7-3 (均匀分布随机变量之和的特征函数)设 X 和 Y 是两个独立分布的随机变量，它们的概率密度函数为

$$f_X(x) = f_Y(x) = \frac{1}{a}\mathrm{rect}\left(\frac{x}{a}\right)$$

试利用特征函数法来计算 $Z \triangleq X + Y$ 的概率密度函数。

解 当然可以用 $f_X(x)$ 卷积 $f_Y(y)$ 来计算 $f_Z(z)$。但是，在这里我们要用特征函数法来计算 $f_Z(z)$，即

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_X(\omega)\Phi_Y(\omega)\mathrm{e}^{-\mathrm{j}\omega z}\mathrm{d}\omega$$

其中

$$\Phi_X(\omega)\Phi_Y(\omega) = \Phi_Z(\omega)$$

因为 X 和 Y 的概率密度函数是相同的，因此可以写成

$$\begin{aligned}\Phi(\omega) &\triangleq \Phi_X(\omega) = \Phi_Y(\omega)\\ &= \frac{1}{a}\int_{-a/2}^{a/2}\mathrm{e}^{\mathrm{j}\omega x}\,\mathrm{d}x\\ &= \frac{\sin(a\omega/2)}{a\omega/2}\end{aligned}$$

所以有

$$\Phi_Z(\omega) = \left(\frac{\sin(a\omega/2)}{a\omega/2}\right)^2 \tag{4.7-8}$$

以及

$$\begin{aligned}f_Z(z) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_Z(\omega)\mathrm{e}^{-\mathrm{j}\omega z}\mathrm{d}\omega\\ &= \frac{1}{a}\left(1-\frac{|z|}{a}\right)\mathrm{rect}\left(\frac{z}{2a}\right)\end{aligned} \tag{4.7-9}$$

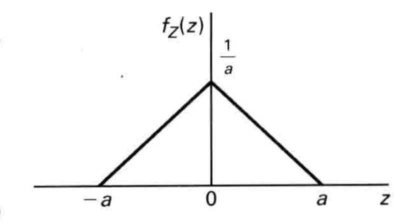


图 4.7-1 $Z = X + Y$ 的概率密度函数，其中随机变量 X 和 Y 是独立的，都服从 $(-a/2, a/2)$ 区间的均匀分布

结果如图 4.7-1 所示。还有一种更为简便的方法来得到式(4.7-9)的结果，那就是查表来计算式(4.7-8)的傅里叶(逆)变换。

对于矩量生成函数的情况，我们可以对特征函数求微分来计算各阶矩，当然前提是这些矩都是存在的。如果将 $\exp(\mathrm{j}\omega X)$ 进行幂级数展开并取数学期望，可得如下结果：

$$\begin{aligned}\Phi_X(\omega) &= E[\mathrm{e}^{\mathrm{j}\omega X}] \\ &= \sum_{n=0}^{\infty} \frac{(\mathrm{j}\omega)^n}{n!} m_n \end{aligned} \tag{4.7-10}$$

从式(4.7-10)很容易得到

$$m_n = \frac{1}{j^n} \Phi_X^{(n)}(0) \tag{4.7-11}$$

其中我们用到了如下的表达符号

$$\Phi_N^{(n)}(0) \triangleq \left. \frac{\mathrm{d}^n}{\mathrm{d}\omega^n} \Phi_X(\omega) \right|_{\omega=0}$$

例 4.7-4　(矩的计算)试计算 $Y = \sin\Theta$ 的前几阶矩，其中 Θ: $U[0, 2\pi]$。

解　可以利用式(4.1-9)的结论。如果 $Y = g(X)$，那么有

$$\mu_Y = \int_{-\infty}^{\infty} y f_Y(y) \mathrm{d}y = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x$$

因此有

$$\begin{aligned} E[\mathrm{e}^{\mathrm{j}\omega Y}] &= \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{j}\omega y} f_Y(y) \mathrm{d}y \\ &= \frac{1}{2\pi} \int_0^{2\pi} \mathrm{e}^{\mathrm{j}\omega \sin\theta} \mathrm{d}\theta \\ &= J_0(\omega) \end{aligned}$$

其中 $J_0(\omega)$ 是第一类贝塞尔函数的第零阶。$J_0(\omega)$ 的幂级数展开结果为

$$J_0(\omega) = 1 - \left(\frac{\omega}{2}\right)^2 + \frac{1}{2!2!}\left(\frac{\omega}{2}\right)^4 - \cdots$$

可知所有的奇数阶矩都是 0。利用式(4.7-11)还可以计算出如下结果：

$$E[Y^2] = m_2 \triangleq (-1)\Phi_X^{(2)}(0) = \tfrac{1}{2}$$

$$E[Y^4] = m_4 = (+1)\Phi_X^{(4)}(0) = \tfrac{3}{8}$$

例 4.7-5　(独立二项分布随机变量之和)设 X 和 Y 是独立同分布的二项分布随机变量，分布参数分布为 n 和 p，概率质量函数为

$$P_X(k) = P_Y(k) = \binom{n}{k} p^k q^{n-k}$$

试计算 $Z = X + Y$ 的概率质量函数。

解　因为 X 和 Y 的取值只能是非负的整数，因此 Z 必然也是非负的整数。我们利用这些方法来解决这个问题：(1)概率密度函数的卷积，其中涉及冲激函数；(2)概率质量函数的离散卷积；(3)特征函数，其中这里所指的离散卷积具体如下：

$$\begin{aligned} P_Z(k) &= \sum_i P_X(i) P_Y(k-i) \\ &= \sum_i \binom{n}{i} p^i q^{n-i} \binom{n}{k-i} p^{k-i} q^{n-(k-i)} \\ &= p^k q^{2n-k} \sum_i \binom{n}{i}\binom{n}{k-i}, \quad k = 0, 1, \cdots, 2n \end{aligned}$$

不过在这里有些麻烦，那就是二项式系数乘积之和的闭式解并不能立即得到①。通过特征函数来计算 Z 的概率质量函数非常容易。首先

$$\begin{aligned}\Phi_X(\omega)=\Phi_Y(\omega)&=\sum_{k=0}^{n}\mathrm{e}^{\mathrm{j}\omega k}\binom{n}{k}p^kq^{n-k}\\&=(p\mathrm{e}^{\mathrm{j}\omega}+q)^n\end{aligned}$$

然后，根据 X 和 Y 相互独立的特性，我们可以得到特征函数。

$$\begin{aligned}\Phi_Z(\omega)&=E[\exp\mathrm{j}\omega(X+Y)]\\&=E[\exp(\mathrm{j}\omega X)]E[\exp(\mathrm{j}\omega Y)]\\&=(p\mathrm{e}^{\mathrm{j}\omega}+q)^{2n}\end{aligned}$$

因此，Z 也服从二项分布，分布参数为 $2n$ 和 p，也就是

$$P_Z(k)=\binom{2n}{k}p^kq^{2n-k},\quad k=0,\cdots,2n$$

通过特征函数来计算得到 $P_Z(k)$，那么就可以计算前面的乘积结果为

$$\binom{2n}{k}=\sum_{i=0}^{n}\binom{n}{i}\binom{n}{k-i}$$

还可以进一步推广这个例题：如果 $X_1,X_2,\cdots,X_N$ 是独立同分布的二项分布随机变量，分布参数为 n 和 p。那么 $Z=\sum_{i=1}^{N}X_i$ 也服从二项分布，分布参数为 Nn 和 p。不管 N 有多大，Z 始终是一个离散的二项分布随机变量②。

例 4.7-6 （泊松分布的方差）请计算泊松分布随机变量的特征函数，再用得到的结果计算方差。设 K 是泊松分布随机变量，其概率质量函数为

$$P_K(k)=\frac{\alpha^k}{k!}\mathrm{e}^{-\alpha}u(k),\quad \alpha>0$$

特征函数表达式如下：

$$\begin{aligned}\Phi_K(\omega)&=\sum_{k=0}^{\infty}\frac{\alpha^k}{k!}\mathrm{e}^{-\alpha}\mathrm{e}^{\mathrm{j}\omega k}\\&=\sum_{k=0}^{\infty}\frac{\left(\alpha\mathrm{e}^{\mathrm{j}\omega}\right)^k}{k!}\mathrm{e}^{-\alpha}\\&=\exp\left[\alpha\left(\mathrm{e}^{\mathrm{j}\omega}-1\right)\right]\end{aligned}$$

可知 $m_2=E[K^2]=\frac{1}{\mathrm{j}^2}\Phi_K^{(2)}(0)=-\Phi_K^{(2)}(0)$。通过取相应的导数可得

$$\Phi_K^{(1)}(\omega)=\Phi_K(\omega)\alpha\mathrm{j}\mathrm{e}^{+\mathrm{j}\omega}$$

以及

$$\begin{aligned}\Phi_K^{(2)}(\omega)&=\Phi_K(\omega)\alpha\mathrm{j}^2\mathrm{e}^{+\mathrm{j}\omega}+\Phi_K^{(1)}(\omega)\alpha\mathrm{j}\mathrm{e}^{+\mathrm{j}\omega}\\&=-\Phi_K(\omega)\alpha\mathrm{e}^{+\mathrm{j}\omega}+\Phi_K(\omega)\left(\alpha\mathrm{j}\mathrm{e}^{+\mathrm{j}\omega}\right)^2\end{aligned}$$

① 我们在第 3 章的例 3.3-9 中已经遇到过这个问题。

② 在讲解中心极限定理的时候，还会提及这个结论。

所以有 $\Phi_K^{(2)}(0)=-1\times\alpha-1\times\alpha^2$，以及 $\mu_2=\alpha+\alpha^2$。因为数学期望 $\mu=a$，所以方差为

$$\begin{aligned}\sigma^2&=m_2-\mu^2\\&=\alpha+\alpha^2-\alpha^2\\&=\alpha\end{aligned}$$

因此泊松分布的方差等于其数学期望。

由于泊松分布的方差等于其数学期望，因此其标准差就是数学期望的平方根。因此对于比较大的数学期望，其分布会相对集中在均值附近。从另一个角度理解，泊松分布与正态分布并不相同，前者的分布参数不是独立的，也就是这些参数并不能随便设置。

例 4.7-7　（一个公平的游戏）有一个叫做“三个玩家，六次击球”的博彩游戏，玩法是这样的：一个玩家任意选择三个棒球选手，如果这三个选手击中球的次数累计等于或大于六次，那么该玩家获胜。存在许多种获胜的组合，比如 A 选手没有击中一次球，B 选手总共击中了三次，而 C 选手也总共击中了三次。这三个选手可以是一个球队的，也可以不是一个球队里的。该博彩游戏的赔率相同，如果玩家赢了，那么他押 1 美元就可以收回 2 美元。试分析这个游戏“公平”吗？也就是说，获胜的概率接近一半吗？

解　设 X_1，X_2，X_3 分别代表选手 A，B，C 击中球的次数。很显然，X_1，X_2 和 X_3 服从独立的二项分布，击中球的总次数为 $Y=\sum_{i=1}^{3}X_i$，在这里我们需要计算 $P[Y\geqslant 6]$。为了简化这个问题，我们假设每个选手每次游戏击球五次，他们的平均击球率都是相同的，假设为 300（对那些不太熟悉棒球术语的读者，这个是指击中球的概率为 0.3）。根据例 4.7-5 的结论可知，Y 服从二项分布，其分布参数为 $n=15$，$p=0.3$。因此有

$$\begin{aligned}P[Y\geqslant 6]&=\sum_{i=6}^{i=15}\binom{15}{i}(0.3)^i(0.7)^{15-i}\\&\approx \operatorname{erf}(6.76)-\operatorname{erf}(0.56)\\&\approx 0.29\end{aligned}$$

在这里我们用到了 1.11 节的一个结论，即用高斯分布来近似二项分布。可以看出，玩家获胜的可能性不足 1/3。如果不考虑这个可怜的赔率，其实可以修改一下这个游戏使得对玩家更“公平”一些。定义一个随机变量 G 为玩家的收益，也定义预期收益为 0 的游戏为公平游戏。如果一个玩家每次获胜能净收入 2.45 美元而不是 1 美元，那么我们将会发现 $E[G]=\$2.45\times 0.29-\$1\times 0.71\approx 0$。当然，如果 $E[G]>0$，那么这个游戏就对玩家有利。有些人就是根据这个准则来玩国家彩票的。

联合特征函数

根据联合矩量生成函数，我们可定义联合特征函数如下：

$$\Phi_{X_1\cdots X_N}(\omega_1,\omega_2,\cdots,\omega_N)=E\left[\exp\left(\mathrm{j}\sum_{i=1}^{N}\omega_iX_i\right)\right]\tag{4.7-12}$$

根据傅里叶变换的特性可知，联合概率密度函数就是 $\Phi_{X_1\cdots X_N}(\omega_1,\cdots,\omega_N)$ 的傅里叶逆变换（符号相反），因此有

$$f_{X_1\cdots X_N}(x_1,\cdots,x_N)=\frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\Phi_{X_1\cdots X_N}(\omega_1,\cdots,\omega_N)\times\exp\left(-\mathrm{j}\sum_{i=1}^{N}\omega_i x_i\right)\mathrm{d}\omega_1\,\mathrm{d}\omega_2\cdots\mathrm{d}\omega_N \tag{4.7-13}$$

我们可以通过求导来计算各阶矩。例如，假设 X 和 Y 是任意两个随机变量($N=2$)，那么有

$$m_{rk}\triangleq E[X^rY^k]=(-\mathrm{j})^{r+k}\Phi_{XY}^{(r,k)}(0,0) \tag{4.7-14}$$

其中

$$\Phi_{XY}^{(rk)}(0,0)\triangleq\left.\frac{\partial^{r+k}\Phi_{XY}(\omega_1,\omega_2)}{\partial\omega_1^r\partial\omega_2^k}\right|_{\omega_1=\omega_2=0} \tag{4.7-15}$$

最后对于离散随机变量，可以用联合概率质量函数来定义它们的联合特征函数。对于 X 和 Y 两个随机变量，可得

$$\Phi_{XY}(\omega_1,\omega_2)\triangleq\sum_i\sum_j \mathrm{e}^{\mathrm{j}(\omega_1x_i+\omega_2y_j)}P_{XY}(x_i,y_j) \tag{4.7-16}$$

例 4.7-8 （独立同分布高斯随机变量的联合特征函数）试计算 X 和 Y 的联合特征函数，其联合概率密度函数如下：

$$f_{XY}=\frac{1}{2\pi}\exp\left[-\frac{1}{2}(x^2+y^2)\right]$$

解 根据式(4.7-12)的定义可知

$$\Phi_{XY}(\omega_1,\omega_2)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}(x^2+y^2)}\mathrm{e}^{\mathrm{j}\omega_1x+\mathrm{j}\omega_2y}\,\mathrm{d}x\,\mathrm{d}y$$

对变量 x 和 y 都采用完全平方法可得

$$\begin{aligned}\Phi_{XY}(\omega_1,\omega_2)&=\mathrm{e}^{-\frac{1}{2}(\omega_1^2+\omega_2^2)}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}[x^2-2\mathrm{j}\omega_1x+(\mathrm{j}\omega_1)^2]}\frac{\mathrm{d}x}{\sqrt{2\pi}}\\&\quad\times\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}[y^2-2\mathrm{j}\omega_2y+(\mathrm{j}\omega_2)^2]}\frac{\mathrm{d}y}{\sqrt{2\pi}}\\&=\mathrm{e}^{-\frac{1}{2}(\omega_1^2+\omega_2^2)}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}(x-\mathrm{j}\omega_1)^2}\frac{\mathrm{d}x}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}(y-\mathrm{j}\omega_2)^2}\frac{\mathrm{d}y}{\sqrt{2\pi}}\\&=\mathrm{e}^{-\frac{1}{2}(\omega_1^2+\omega_2^2)}\end{aligned}$$

因为单位方差的高斯曲线下的面积为 1，因此最后一步成立。

例 4.7-9 （两个离散随机变量的联合特征函数）试计算离散随机变量 X 和 Y 的联合特征函数，其联合概率质量函数为

$$P_{XY}(k,l)=\begin{cases}\frac{1}{3}, & k=l=0\\ \frac{1}{6}, & k=\pm1,l=0\\ \frac{1}{6}, & k=l=\pm1\\ 0, & \text{其他}\end{cases}$$

解 根据式(4.7-16)可得

$$\Phi_{XY}(\omega_1,\omega_2)=\sum_{k=-1}^{1}\sum_{l=-1}^{1}\mathrm{e}^{\mathrm{j}(\omega_1 k+\omega_2 l)}P_{XY}(k,l)$$
$$=\frac{1}{3}+\frac{1}{3}\cos\omega_1+\frac{1}{3}\cos(\omega_1+\omega_2)$$

因为 $\mu_X=\mu_Y=0$，根据式(4.7-14)和式(4.7-15)可得

$$\sigma_X^2\triangleq m_{20}=-(-\mathrm{j})^2[\cos\omega_1+\cos(\omega_1+\omega_2)]\left.\frac{1}{3}\right|_{\omega_1=\omega_2=0}=\frac{2}{3}$$
$$\sigma_Y^2\triangleq m_{02}=-(-\mathrm{j})^2\left.\frac{1}{3}\cos(\omega_1+\omega_2)\right|_{\omega_1=\omega_2=0}=\frac{1}{3}$$
$$m_{11}=-(-\mathrm{j})^2\left.\frac{1}{3}\cos(\omega_1+\omega_2)\right|_{\omega_1=\omega_0=0}=\frac{1}{3}$$

此外，相关系数 ρ 可计算如下：

$$\rho=\frac{m_{11}}{\sigma_X\sigma_Y}=\frac{\frac{1}{3}}{\sqrt{\frac{2}{3}}\sqrt{\frac{1}{3}}}=\frac{1}{\sqrt{2}}=0.707$$

例 4.7-10　(相关高斯随机变量的联合特征函数)另外一个例子，请计算 X 和 Y 的联合特征函数，其联合概率密度函数为

$$f_{XY}(x,y)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right)$$

为求解这个题目，我们需要用到两个结论：

(1)高斯分布随机变量 Z，其均值为 0，方差为 σ_Z^2，那么它的特征函数为

$$E[\mathrm{e}^{\mathrm{j}\omega Z}]=\exp\left[-\frac{1}{2}\sigma_Z^2\omega^2\right]\tag{4.7-17}$$

特别地，当 $\omega=1$ 时

$$E[\mathrm{e}^{\mathrm{j}Z}]=\exp\left[-\frac{1}{2}\sigma_Z^2\right]\tag{4.7-18}$$

结论(1)的证明　可根据特征函数的定义，以及 $f_Z(z)=(2\pi\sigma_Z^2)^{-1/2}\exp\left(-\frac{1}{2}\frac{z^2}{\sigma_Z^2}\right)$，再用附录 A 中介绍的完全平方法即可。

(2)如果 X 和 Y 是零均值的联合高斯分布随机变量，那么对于任意的实数 ω_1 和 ω_2，定义如下的随机变量：

$$Z\triangleq\omega_1 X+\omega_2 Y$$
$$W\triangleq X$$

上述变量是联合高斯的，而且 Z 的边缘密度函数也是高斯的。

结论(2)的证明　根据式(3.4-11)或式(3.4-12)便可计算得到 $f_{ZW}(z,w)$。很容易发现 Z 和 W 是联合高斯的，因此 Z 的边缘概率密度函数也是高斯的，并且 $\bar{Z}=0$。Z 的方差可计算如下：

$$\begin{aligned}\mathrm{Var}(Z) &= E[(\omega_1 X + \omega_2 Y)^2] \\ &= \omega_1^2 \mathrm{Var}[X] + \omega_2^2 \mathrm{Var}[Y] + 2\omega_1\omega_2 \overline{XY}\end{aligned}$$

根据 $\sigma_X^2 = \sigma_Y^2 \triangleq 1$ 可得 $\sigma_Z^2 = \omega_1^2 + \omega_2^2 + 2\omega_1\omega_2\rho$。

最后，对于 $Z = \omega_1 X + \omega_2 Y$，根据式(4.7-18)可得

$$E[\mathrm{e}^{\mathrm{j}(\omega_1 X + \omega_2 Y)}] = \mathrm{e}^{-\frac{1}{2}(\omega_1^2 + \omega_2^2 + 2\omega_1\omega_2\rho)} \tag{4.7-19}$$

式(4.7-19)就是两个零均值，单位方差的相关高斯随机变量的联合特征函数。当 $\rho = 0$ 时，随机变量变得不相关，二者此时是独立的，在例 4.7-8 中也有这个结论。

可直接将上述结论扩展到多个离散随机变量的情况。如果不采用矩阵，此时的符号表达就会有些麻烦。

中心极限定理

经常会说到“大量随机变量之和趋向于正态分布”，这个结论在什么情况下成立？中心极限定理就将阐述这个重要的观点。

中心极限定理①一般是指：大量独立随机变量之和经归一化后趋向于正态分布，这些随机变量 $X_1, \cdots, X_n$ 的均值都为零，方差 $\sigma_1^2, \cdots, \sigma_n^2$ 都是有限值，同时各个独立的方差 σ_k^2 与 $\sum_{i=1}^{n} \sigma_i^2$ 相比都很小，其中 $k = 1, \cdots, n$。对方差的这个限制就是所谓的 Lindeberg 条件，在 Feller[4-1, p.262] 中有详细的讨论。接下来我们将给出中心极限定理的一般形式，并对一个特例给出证明。

定理 4.7-1 设 $X_1, \cdots, X_n$ 为 n 个相互独立的随机变量(标量)，概率分布函数分别为 $F_{X_1}(x_1), F_{X_2}(x_2), \cdots, F_{X_n}(x_n)$，均值和方差为

$$\mu_{X_k} = 0, \qquad \mathrm{Var}[X_k] = \sigma_k^2$$

设

$$s_n^2 \triangleq \sigma_1^2 + \cdots + \sigma_n^2$$

如果对于一个给定的 $\varepsilon > 0$，当 n 足够大时 σ_k 满足

$$\sigma_k < \varepsilon s_n, \qquad k = 1, \cdots, n$$

那么下面的归一化和就收敛于标准正态分布。

$$Z_n \triangleq (X_1 + \cdots + X_n)/s_n$$

在这里用 $1/2 + \mathrm{erf}(z)$ 来表示，也即是 $\lim\limits_{n\to\infty} F_{Z_n}(z) = 1/2 + \mathrm{erf}(z)$。这称为分布的收敛性。

分布的收敛性在接下来的章节会讨论。下面我们给出前面定理一个特例的证明。

定理 4.7-2 设 $X_1, X_2, \cdots, X_n$ 是独立同分布的随机变量，其中 $\mu_{X_i} = 0$，$\mathrm{Var}[X_i] = 1$，$i = 1, \cdots, n$，那么下面的表达式趋向于正态分布。

$$Z_n \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

---

① Abraham De Moivre 针对伯努利随机变量这个特例，于1733 年最早证明该定理。该定理更一般化的证明由 J. W. Lindeberg 证明，参见 *Mathematische Zeitschrift*, vol. 15 (1922), pp. 211-225。

同时它的特征函数 Φ_{Z_n} 满足下面的条件：

$$\lim_{n\to\infty} \Phi_{Z_n}(\omega) = \mathrm{e}^{-\frac{1}{2}\omega^2}$$

这其实就是服从 $N(0, 1)$ 分布随机变量的特征函数。

证明　设 $W_i \triangleq X_i/\sqrt{n}$，令 $\Phi_{X_i}(\omega)$ 和 $f_{X_i}(x)$ 分布表示 X 的特征函数和概率密度函数，那么有

$$\begin{aligned}\Phi_{W_i} &\triangleq E[\mathrm{e}^{\mathrm{j}\omega W_i}] \\ &= E[\mathrm{e}^{\mathrm{j}(\omega/\sqrt{n})X_i}] \\ &= \Phi_{X_i}\left(\frac{\omega}{\sqrt{n}}\right)\end{aligned}$$

因为 $\Phi_{X_i}(\omega)$ 和 $\Phi_{W_i}(\omega)$ 并不取决于变量 i，因此我们可以写成 $\Phi_{X_i}(\omega) \triangleq \Phi_X(\omega)$，$\Phi_{W_i}(\omega) \triangleq \Phi_W(\omega)$。根据微积分知识可知，任何一个函数 $\Phi(\omega)$ 如果在 ω_0 附近导数存在的话，那么该函数可展开成泰勒级数

$$\Phi(\omega) = \sum_{l=0}^{\infty} \frac{1}{l!}\Phi^{(l)}(\omega_0)(\omega-\omega_0)^l$$

其中 $\Phi^{(l)}(\omega_0)$ 表示 $\Phi(\omega)$ 在 ω_0 处的第 l 阶导数。进一步地，如果在区间 $[\omega_0, \omega]$ 里导数都是连续的，那么 $\Phi(\omega)$ 就可以用一个有限长度的泰勒级数表示，也即是

$$\Phi(\omega) = \sum_{l=0}^{L-1} \frac{1}{l!}\Phi^{l}(\omega_0)(\omega-\omega_0)^l + A_L(\omega)$$

其中因子 $A_L(\omega)$ 表达式为

$$A_L(\omega) \triangleq \frac{1}{L!}\Phi^L(\zeta)(\omega-\omega_0)^L$$

ξ 是区间 $[\omega_0, \omega]$ 中的某个点。将这个结果应用于 $\Phi_W(\omega)$，其中 $\omega_0 = 0$，那么有

$$\begin{aligned}&\Phi_W(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{j}\omega x/\sqrt{n}} f_X(x)\,\mathrm{d}x \\ &\Phi_W^{(0)}(0) = 1 \\ &\Phi_W^{(1)}(0) = \int_{-\infty}^{\infty} \mathrm{j}\,\frac{x}{\sqrt{n}}\mathrm{e}^{\mathrm{j}\omega x/\sqrt{n}} f_X(x)\,\mathrm{d}x\bigg|_{\omega=0} = 0 \\ &\Phi_W^{(2)}(0) = \int_{-\infty}^{\infty} \left(\frac{\mathrm{j}x}{\sqrt{n}}\right)^2 \mathrm{e}^{\mathrm{j}\omega x/\sqrt{n}} f_X(x)\,\mathrm{d}x\bigg|_{\omega=0} = -\frac{1}{n}\end{aligned}$$

因此可得

$$\Phi_W(\omega) = 1 - \frac{1}{2n}\omega^2 + \frac{R_2'(\omega)}{n\sqrt{n}}$$

其中

$$R_2'(\omega) \triangleq -\mathrm{j}\omega^3 \int_{-\infty}^{\infty} x^3 \mathrm{e}^{\mathrm{j}\zeta x/\sqrt{n}} f_X(x)\,\mathrm{d}x/6$$

因为 $Z_n = \sum_{i=1}^{n} W_i$，我们可以得到

$$\Phi_{Z_n}(\omega)=[\Phi_W(\omega)]^n$$

或者

$$\ln\Phi_{Z_n}(\omega)=n\ln\Phi_W(\omega)$$

现在我们将用到一个结论，对于任意的 h，只要满足 $|h|<1$，就有

$$\ln(1+h)=h-\frac{h^2}{2}+\frac{h^3}{3}-\cdots$$

对于任意固定的 ω，可以选择一个足够大的 n 使得下面表达式成立(设 $R_2'\triangleq R_2'(\omega)$)：

$$\left|-\frac{\omega^2}{2n}+\frac{R_2'}{n\sqrt{n}}\right|<1$$

根据上面的结论

$$\begin{aligned}\ln\Phi_{Z_n}(\omega)&=n\ln\left[1-\frac{\omega^2}{2n}+\frac{R_2'}{n\sqrt{n}}\right]\\&\approx n\left[-\frac{\omega^2}{2n}+\frac{R_2'}{n\sqrt{n}}-\frac{1}{2}\left(-\frac{\omega^2}{2n}+\frac{R_2'}{n\sqrt{n}}\right)^2+\frac{1}{3}\left(-\frac{\omega^2}{2n}+\frac{R_2'}{n\sqrt{n}}\right)^3+\cdots\right]\\&=-\frac{\omega^2}{2}+\text{涉及的因子}\ n^{-1/2},n^{-1},n^{-3/2},\cdots.\end{aligned}$$

因此有

$$\lim_{n\to\infty}[\ln\Phi_{Z_n}(\omega)]=-\frac{\omega^2}{2}$$

其等效表达式为

$$\lim_{n\to\infty}\Phi_{Z_n}(\omega)=\mathrm{e}^{-\omega^2/2}$$

上式就是 $N(0,1)$ 分布随机变量的特征函数。如果 $\lim\limits_{n\to\infty}f_{Z_n}(z)$ 是正态分布的概率密度函数，那么下面的等式就应该成立：

$$\begin{aligned}\lim_{n\to\infty}\Phi_{Z_n}(\omega)&\triangleq\lim_{n\to\infty}\left(\int_{-\infty}^{\infty}f_{Z_n}(z)\mathrm{e}^{\mathrm{j}\omega z}\mathrm{d}z\right)\\&\overset{?}{=}\int_{-\infty}^{\infty}\left(\lim_{n\to\infty}f_{Z_n}(z)\right)\mathrm{e}^{\mathrm{j}\omega z}\mathrm{d}z\end{aligned}$$

但是，求极限运算和积分运算并不是总能交换顺序的，因此我们不能认为 Z_n 的概率密度函数就收敛于 $N(0,1)$。实际上，从例 4.7-5 知道 n 个独立同分布的二项分布随机变量之和仍然服从二项分布，而且 n 多大都无所谓。此外，二项分布的概率密度(质量)函数是非连续的，而高斯分布的概率密度函数是连续的，不管 n 有多大这个事实也是无法改变的。但是，当 n 比较大时，二项分布概率密度函数的积分与高斯概率密度函数的积分很相像。这就是为什么说 Z_n 的概率密度函数趋向于高斯分布，而不说等于高斯分布。

聪明的读者可能会发现在前面的推导过程中，我们只证明了特征函数收敛于正态分布，而没有证明概率分布函数收敛于正态分布。为了证明后面这个结论，我们需要用到连续性定理①，该定理具体为：对于随机序列 Z_i，其特征函数和概率分布函数分别为 $\Phi_i(\omega)$ 和 $F_i(z)$，

① 参见 Feller[4-1，p. 508]。

$i=1,\cdots,n$，此外 $\Phi(\omega)\triangleq\lim\limits_{n\to\infty}\Phi_n(\omega)$，$\Phi(\omega)$ 在 $\omega=0$ 处连续，那么有 $F(z)=\lim\limits_{n\to\infty}F_n(z)$。

例 4.7-11　（中心极限定理的应用）设 X_i 是一个独立同分布的随机序列，$i=1,\cdots,n$，$E[X_i]=\mu_X$，$\text{Var}[X_i]=\sigma_X^2$。设 $Y\triangleq\sum_{i=1}^{n}X_i$，其中 n 足够大。试根据中心极限定理来计算 $P[a<Y\leqslant b]$，其中 $Z\triangleq(Y-E[Y])/\sigma_Y$，$\sigma_Y>0$。

$$P[a<Y\leqslant b]=P[a'<Z\leqslant b']$$

其中

$$a'\triangleq\frac{a-E[Y]}{\sigma_Y}$$

$$b'=\frac{b-E[Y]}{\sigma_Y}$$

以及有

$$\sigma_Y=\sqrt{n}\sigma_X$$

注意到 Z 是一个零均值，方差为 1 的高斯分布随机变量，在这里用到了大量独立同分布随机变量之和。实际上，对 Z 适当进行一下数学变换可得

$$Z=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{X_i-\mu_X}{\sigma_X}\right)$$

因此有下面的表达式成立

$$P[a'<Z\leqslant b']\approx\frac{1}{\sqrt{2\pi}}\int_{a'}^{b'}\mathrm{e}^{-\frac{1}{2}z^2}\mathrm{d}z$$

把中心极限定理称为“正态收敛性定理”可能更合适一些，不过在中心极限定理中“中心”这个词语是为了强调概率分布函数与正态概率分布函数在其中心附近收敛，也即是在其均值附近收敛。而实际上，尽管所有的概率分布函数都会在 $\pm\infty$ 处收敛，在首尾两端中心极限定理给出的估计值都非常糟糕，除非这个值本来就很小。在习题 4.59 中会详细描述这个现象。

有一种基于计算机的工程分析方法称为蒙特卡罗仿真，经常需要用到随机数。有许多软件可以用来产生随机数，不过这些产生的随机数并不是真正随机的。因为这些随机数都是通过某种确定性的算法产生的，这意味着任何一个人利用该算法，都可以产生同样的随机数。这种随机数称为伪随机数，如果仿真并不需要很多的随机数，那么这些随机数对这个仿真实验就是充分的。而对于那些需要很多随机数的仿真实验而言，比如模拟原子过程，此时就很不容易找到一个能产生充分数量的随机数发生器。很多情况下产生的随机数都是可重复的，也就是说这些随机数是周期的，高度相关的，或者存在某种程度的偏差。此外，如果采用一个天然的随机数发生器（比如 X 射线源中发射光子的数目，或者光电探测器在电子带中探测到的光电子数目）也会遇到一些麻烦，此时我们并不能确切知道到底是什么规律在起作用。而且即使我们确切知道了是什么规律，光子或光电子的行为也会影响随机数的分布。

我们假设在大多数的个人计算机软件平台上都可以运行均匀分布随机数发生器（URNG），而且产生的均匀分布随机数 X 是无偏的。如果假设这些随机数服从 $(0,1)$ 区间的均匀分布，那么接下来的问题就是：如何把这些产生的均匀分布随机数转换为高斯分布的随机数？我们可以用中心极限定理来实现这个目的，具体步骤为：假设 X_i 是这个均匀分布随机数发生器产生的第 i 个随机数，那么当 n 足够大的时候（比如 $n>10$），Z 将近似服从高斯分布。

$$Z = X_1 + \cdots + X_n$$

Z 的概率密度函数其实是 n 个单位冲激的卷积，卷积的结果非常类似于高斯分布。但在首尾两端二者并不相同，这是因为在这里 $0 \leqslant Z \leqslant n$，而如果 Z 是真正的高斯分布随机变量的话，其分布区间应该是 $-\infty < Z < \infty$。

4.8 一些补充例子

例 4.8-1 设 X_i 是 n 个独立同分布的伯努利随机变量，$i = 1, \cdots, n$，其概率质量函数为

$$P_{X_i}(x) = \begin{cases} p^x(1-p)^{1-x}, x = 0, 1 \\ 0, \text{其他} \end{cases}$$

试证明 $Z = \sum_{i=1}^{n} X_i$ 服从二项分布，且概率质量函数为 $b(k; n, p) \triangleq \binom{n}{k} p^k q^{n-k}$。

解 伯努利分布的特征函数为 $\Phi_{X_i}(\omega) = \sum_{x=0}^{1} \mathrm{e}^{\mathrm{j}\omega x} p^x (1-p)^{1-x} = p\mathrm{e}^{\mathrm{j}\omega} + q$，其中 $q = 1 - p$。根据式(4.7-4)的结论可知

$$\Phi_Z(\omega) = \prod_{i=1}^{n} (p\mathrm{e}^{\mathrm{j}\omega} + 1) = (p\mathrm{e}^{\mathrm{j}\omega} + 1)^n$$

得到的结果与例 4.7-5 中二项分布随机变量的特征函数是一致的。

例 4.8-2 设 Z 服从二项分布，概率质量函数为 $b(k; n, p) \triangleq \binom{n}{k} p^k q^{n-k}$，其中 $n \gg 1$，对于事件 $\{a \leqslant Z \leqslant b\}$，试用中心极限定理计算 $P[a \leqslant Z \leqslant b]$，其中 a 和 b 是常数。

解 从例 4.8-1 可知，Z 可看成 n 个独立同分布伯努利随机变量之和，因此我们可以将 Z 写为 $Z = X_1 + \cdots + X_n$，其中 $E[Z] = np$，$\mathrm{Var}[Z] = npq \gg pq$，$n$ 是一个比较大的数。现在的问题就是如何将中心极限定理用到这里。事件 $\{a \leqslant Z \leqslant b\}$ 与事件 $\left\{\frac{a - np}{\sqrt{npq}} \leqslant \frac{Z - np}{\sqrt{npq}} \leqslant \frac{b - np}{\sqrt{npq}}\right\}$ 是等价的。通过变量替换 $a' \triangleq \frac{a - np}{\sqrt{npq}}$，$b' \triangleq \frac{b - np}{\sqrt{npq}}$ 以及 $Z' \triangleq \frac{Z - np}{\sqrt{npq}}$，又可以写成 $\{a' \leqslant Z' \leqslant b'\}$，此时 Z' 是一个零均值，单位方差的随机变量。利用例 4.7-11 的结论，以及根据中心极限定理可以得到

$$P[a \leqslant Z \leqslant b] \approx \frac{1}{\sqrt{2\pi}} \int_{a'}^{b'} \exp[-\frac{1}{2}z^2] \mathrm{d}z$$

根据式(1.11-3)中定义的标准正态分布函数 $F_{\mathrm{SN}}(x)$ 可知，上述结果又可以写成如下形式

$$P[a \leqslant Z \leqslant b] \approx F_{\mathrm{SN}}\left[\frac{b - np}{\sqrt{npq}}\right] - F_{\mathrm{SN}}\left[\frac{a - np}{\sqrt{npq}}\right]$$

当 $n \gg 1$ 时，式(1.11-5)中的纠正因子 0.5 可以忽略不计。

例 4.8-3 设 Z 是二项分布随机变量，均值为 np，标准差为 $\sqrt{npq}$。试根据中心极限定理中有关正态近似的结论来计算以下事件的概率：$\{np - \sqrt{npq} \leqslant Z \leqslant np + \sqrt{npq}\}$，$\{np - 2\sqrt{npq} \leqslant Z \leqslant np + 2\sqrt{npq}\}$ 以及 $\{np - 3\sqrt{npq} \leqslant Z \leqslant np + 3\sqrt{npq}\}$。

解 通过变量替换 $Z' \triangleq \dfrac{Z-np}{\sqrt{npq}}$，可以将上面三个事件转变为 $\{-1\leqslant Z'\leqslant 1\}$，$\{-2\leqslant Z'\leqslant 2\}$ 和 $\{-3\leqslant Z'\leqslant 3\}$。通过中心极限定理中有关正态近似的结论可知，随机变量 Z' 均值为零，方差为1，因此可得

$$P[-1\leqslant Z'\leqslant 1]=F_{\mathrm{SN}}(1)-F_{\mathrm{SN}}(-1)\approx 0.683$$
$$P[-2\leqslant Z'\leqslant 2]=F_{\mathrm{SN}}(2)-F_{\mathrm{SN}}(-2)\approx 0.954$$
$$P[-3\leqslant Z'\leqslant 3]=F_{\mathrm{SN}}(3)-F_{\mathrm{SN}}(-3)\approx 0.997$$

上面最后一个事件(几乎)是肯定会发生的。平均而言，在1000次重复实验中只有3次不会发生。

例 4.8-4 设 X_i 是独立同分布的泊松分布随机变量，概率质量函数为 $P[k]=\mathrm{e}^{-2}\dfrac{2^k}{k!}$，$k=0,1,2,\cdots$，$i=1,\cdots,100$。在这里 k 表示在一个给定的时间段内事件发生的次数。设 $Z=\sum_{i=1}^{100}X_i$，并且有 $E[Z]=200$，$\mathrm{Var}[Z]=200$。在多通道接收机中，接收到的数据包之和一般服从这种规律。试根据中心极限定理来计算事件 $\{190\leqslant Z\leqslant 210\}$ 的概率。

解 因为 Z 是大量独立同分布随机变量之和，任何一个分量的方差都远小于各个分量之和的方差，因此可用中心极限定理中关于正态近似的结论来计算这个事件的概率。定义 $Z'=\dfrac{Z-200}{14.14}$ 为零均值，单位方差的随机变量，利用这个变量定义，待计算事件变为 $\{-0.707\leqslant Z'\leqslant 0.707\}$。通过正态近似可知 $F_{\mathrm{SN}}(0.707)-F_{\mathrm{SN}}(-0.707)\approx 0.52$。

小结

在本章中，我们讨论了一个或多个随机变量的各类平均值，以及这些平均值的应用。首先，定义了一个随机变量 X 的均值或数学期望，然后根据 X 的概率质量(密度)函数可直接计算 $Y=g(X)$ 的数学期望。我们还简要讨论了条件数学期望的重要概念，以及如何根据一个随机变量的条件数学期望来计算它的数学期望。此外，还介绍了用单一统计数据(比如 X 的均值 μ_X)不能充分的描述 X 的性质，从这点出发我们介绍了矩的概念，即 X 功率的平均值。我们介绍了根据最大熵原则，如何根据矩来估计概率密度函数，并引出了联合矩的概念。我们也介绍了两个随机变量的协方差可以看成二者关系的一种测量，即根据线性预测模型通过一个变量预测另一个变量。通过一个反例，我们强调随机变量不相关并不意味着二者独立，独立性是一个更强的条件。还介绍了两个随机变量的联合高斯概率密度函数，对于高斯分布而言，不相关和独立是等价的。我们还给读者介绍了一些重要的限和不等式，比如切比雪夫不等式，施瓦茨不等式，以及切尔诺夫限等，并介绍了在概率问题中如何使用它们。

在本章的第二部分主要介绍了矩量生成函数(MGF)，特征函数(CF)和中心极限定理(CLT)。我们证明了一个随机变量的矩量生成函数和特征函数，实际上就是概率密度函数的拉普拉斯变换和傅里叶变换，以及如何通过矩量生成函数或特征函数来计算各阶距，当然前提是这些矩存在。此外还介绍了矩量生成函数和特征函数的一些重要性质，以及如何利用特征函数来计算随机变量之和的概率密度函数。

此外还介绍了中心极限定理，这是概率论中最重要的结论之一，也是许多随机现象为什么

会体现出正态特性的理论基础。中心极限定理指在相对宽松的数学约束条件下，独立随机变量之和的概率分布函数趋向于正态分布。

在本章最后，通过一些补充的例题介绍了中心极限定理的应用。

习题

（带 * 号的问题带有一定的难度，需要花更多精力对教材进一步的学习。）

4.1　试计算下面集合的均值和标准偏差：3.02, 5.61, −2.37, 4.94, −6.25, −1.05, −3.25, 5.81, 2.27, 0.54, 6.11, −2.56。

4.2　请计算 X 的均值 $E[X]$，其中 X 是伯努利随机变量

$$X=\begin{cases}1, & P_X(1)=p>0\\ 0, & P_X(0)=1-p>0\end{cases}$$

4.3　设 $X=a$（a 为一个常数），请证明 $E[X]=a$①。

4.4　设 X 是一个二项分布随机变量

$$P_X(k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,\cdots,n,\quad 0<p<1$$

请计算 $E[X]$。

4.5　设 X 为一个均匀分布随机变量，概率密度函数为

$$f_X(x)=\begin{cases}(b-a)^{-1}, & 0<a<x<b\\ 0, & \text{其他}\end{cases}$$

请计算 $E[X]$。

4.6　设 X 的概率密度函数为 $f_X(x)=\begin{cases}2x & 0<x<1\\ 0, & \text{其他}\end{cases}$

(a) 计算 $F_X(x)$。

(b) 计算 $E[X]$。

(c) 计算 σ_X^2。

4.7　如果 $x=0,1,\cdots,k$ 时 $P_X(x)=\dfrac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}}$，其余情况 $P_X(x)=0$。请计算 $E[X]$。这种概率质量函数称为超几何分布，m, n 和 k 都是正整数。

4.8　对于习题4.5，设 $Y\triangleq X^2$，请利用式(4.1-8)计算 Y 的概率密度函数和数学期望 $E[Y]$，再请利用式(4.1-9)计算 $E[Y]$。

4.9　设 $Y\triangleq X^2+1$，X 的概率密度函数如下，请计算 $E[Y]$ 和 σ_Y^2。

$$f_X(x)=\begin{cases}2x, & 0<x<1\\ 0, & \text{其他}\end{cases}$$

4.10　设 X 是一个泊松分布随机变量，分布参数为 a，如果令 $Y\triangleq X^2+b$，请计算 $E[Y]$。

4.11　请利用均值的积分定义证明：高斯分布随机变量 X: $N(\mu,\sigma^2)$ 的均值为 μ。

4.12　在物理课程中，应该学习过动量 $p=mv$ 的概念，不过书上讲的一般都是确定性或非随机特性的动量。但在实际情况下，对质量 m 和速度 v 的测量不会是完全准确的，也就是说不可避免地有些不确定性因素在其中。在这里，我们把这些参数看成随机变量。假设质量 M 是一个随机参量，其概率密度函数为

---

① 原文是证明 $E[Y]=a$，明显有误——译者注。

$f_M(m)$，速度 V 也是一个随机变量，概率密度函数为 $f_V(v)$。假设数学期望 $\mu_M = E[M]$ 和 $\mu_V = E[V]$ 也是已知的(物理课程中应该讲到了如何测量这些参数)。假设 M 和 V 是独立的，非负的随机变量。

(a)试根据概率密度函数 $f_M(m)$ 和 $f_V(v)$ 写出动量 $P = MV$ 的概率密度函数。

(b)试根据 μ_M 和 μ_V 写出动量的数学期望 $\mu_P = E[P]$。

4.13 如果 X 是一个连续随机变量且 $E[X]$ 存在，请证明 $|E[X]| \leqslant E[|X|]$。当 X 是离散随机变量时，请证明同样的结论。

4.14 如果 $E[g_i(X)]$ 存在，$i = 1, \cdots, N$，请证明下面的结论成立：

$$E\left[\sum_{i=1}^{N} g_i(X)\right] = \sum_{i=1}^{N} E[g_i(X)]$$

4.15 随机对 20 个家庭进行采样，各个家庭拥有的孩子数量为 3，2，0，1，0，0，3，2，5，0，1，1，2，0，1，0，0，0，6，3。(a)总体而言，每个家庭平均拥有多少个孩子？(b)对于至少拥有一个孩子的家庭而言，每个家庭平均拥有多少个孩子？

4.16 设 $B \triangleq \{a < X \leqslant b\}$。如果 X 是一个随机变量，试推导出 $E[X|B]$ 的一般表达式。当 X: $N(0, 1)$，$B = \{-1 < X \leqslant 2\}$ 时，请计算 $E[X|B]$。

4.17 (Papoulis[4-3])设 $Y = h(X)$，我们希望粗略计算 $E[h(X)]$ 和 $E[h^2(X)]$。假设 $h(x)$ 可以被幂级数展开，也即是它的所有导数都存在。进一步假设所有二阶以上的导数都可以忽略不计。已知 $E[X] = \mu$，$\mathrm{Var}(X) = \sigma^2$，请证明下面的结论：

(a) $E[h(X)] \approx h(\mu) + h''(\mu)\sigma^2/2$

(b) $E[h^2(X)] \approx h^2(\mu) + ([h'(\mu)]^2 + h(\mu)h''(\mu))\sigma^2$

4.18 设 X 和 Y 是两个随机变量，X 的概率密度函数 $f_X(x)$ 为

$$f_X(x) = \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & \text{其他} \end{cases}$$

当 $X = x$ 时，Y 的条件概率密度函数 $f_{Y|X}(y|x)$ 为

$$f_{Y|X}(y|x) = \begin{cases} 2y/x^2 & 0 < y < x < 1 \\ 0 & \text{其他} \end{cases}$$

(a)请计算 X 和 Y 的联合概率密度函数 $f_{X,Y}(x. y)$。

(b)请计算 $X = x$ 时 Y 的条件均值，即 $E[Y|x]$。

(c)请计算边缘概率密度函数 $f_Y(y)$ 和均值 $E[Y]$。

4.19 高清电视机(HDTV)的组装可以用一个典型的模型来建模。假设在同一个公司里面有 A，B，C 三个不同的组装车间，因为这三个车间里的工人工作经验不同，因此各个车间组装出来的电视机质量会有差异。假设产品寿命 X 的概率密度函数为(单位：年)

$$f_X(x) = \frac{1}{5}\exp(-x/5)u(x), \quad \text{A}$$

$$f_X(x) = \frac{1}{6.5}\exp(-x/6.5)u(x), \quad \text{B}$$

$$f_X(x) = \frac{1}{10}\exp(-x/10)u(x), \quad \text{C}$$

其中 $u(x)$ 是单位阶跃函数。A 车间的产量是 B 车间的三倍，而 B 车间的产量是 C 车间的两倍。生产出来的所有电视机都会发往一个中心仓库，在那里混装后再运往各个零售商店。对于一台随机购买的电视机，请计算这台电视机的预期寿命。

4.20 一个信号源传输一个信号 Θ，其概率密度函数为

$$f_\Theta(\theta) = \begin{cases} (2\pi)^{-1}, & 0 < \theta \leqslant 2\pi \\ 0, & \text{其他} \end{cases}$$

由于加性高斯白噪声的影响，接收信号 Y 的概率密度函数为(此时 $\Theta = \theta$)

$$f_{Y|\Theta}(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{y-\theta}{\sigma}\right)^2\right]$$

请计算 $E[Y]$。

4.21 请计算 X 的方差，如果 X 服从：(a)伯努利分布；(b)二项分布；(c)泊松分布；(d)高斯分布；(e)瑞利分布。

4.22 有两类互联网服务提供商(ISP)给用户提供分发数据包的服务，不过这种服务会随机失败。我们发现故障时间的分布服从指数分布，分布参数分别为 μ_1 和 μ_2。故障时间是随机变量，我们把它们分别称为 T_1 和 T_2，并且假设它们是相互独立的。有 30% 的服务商是第 1 种类型的，有 70% 的是第 2 种类型的。如果随机选取服务商，并且用随机变量 T 来代表故障时间。

(a)请计算 $E[T]$。

(b)请计算 $E[T^2]$。

(c)请计算标准偏差 σ_T。

4.23 设 X 和 Y 是独立的随机变量，服从分布 $N(0, 1)$。请计算 $Z \triangleq \sqrt{X^2+Y^2}$ 的均值和方差。

4.24 设 Y_1, Y_2, Y_3 是三个独立同分布的标准正态分布随机变量，设 $Y_1 < Y_2 < Y_3$①。

(a)计算 $f_{Y_1Y_2Y_3}(y_1, y_2, y_3)$。

(b)计算 $E[Y_i]$, $i = 1, 2, 3$。

4.25 设 $0 < x < y < 1$ 时 $f_{XY}(x, y) = 2$，其他区间 $f_{XY}(x, y) = 0$，请计算 $E[Y]$ 和 σ_Y^2。

4.26 设 $|\rho| < 1$ 时 $f_{XY}(x, y)$ 的表达式如下：

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(-\frac{x^2+y^2-2\rho xy}{2\sigma^2(1-\rho^2)}\right)$$

请证明 $E[Y] = 0$, $E[Y|X = x] = \rho x$。根据这个结果，可否用 X 的观测值来估计 Y?

4.27 设 X 和 Y 为两个高斯随机变量，均值为 0，方差为 σ^2，设

$$Z \triangleq \frac{1}{2}(X+Y)$$

(a)如果 X 和 Y 是相互独立的，请计算 Z 的均值和方差。

(b)如果 X 和 Y 不是独立的，设 ρ 为 X 和 Y 的相关系数，请再次计算 Z 的均值和方差(提示：你的答案可能与 ρ 有关)。

(c)当 $\rho = -1$, $\rho = 0$ 和 $\rho = +1$ 时会是什么结果？这种情况经常发生吗②?

4.28 对于联合高斯概率密度函数，均值 $\mu_X = \mu_Y = 0$，方差 $\sigma_X = \sigma_Y \triangleq \sigma$，请证明当 $\rho \to 1$ 时联合高斯概率密度函数趋向于下面的表达式：

$$f_{XY}(x,y) \to \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right] \delta(y-x)$$

4.29 对于一个概率空间 $\mathscr{P} = (\Omega, \mathscr{F}, P)$，设 $\Omega = \{\zeta_1, \cdots, \zeta_5\} = \{-1, -\frac{1}{2}, 0, \frac{1}{2}, 1\}$, $P[\{\zeta_i\}] = \frac{1}{5}$, $i = 1, \cdots, 5$。在 $\mathscr{P}$ 上定义两个随机变量如下：

$$X(\zeta) \triangleq \zeta \qquad 和 \qquad Y(\zeta) \triangleq \zeta^2$$

(a)请证明 X 和 Y 是独立的随机变量。

(b)请证明 X 和 Y 是不相关的。

4.30 两个随机变量 X 和 Y 的条件高斯密度函数定义如下：

---

① 习题 4.24 原文有两处错误，第一是说的是 X_1, X_2 和 X_3 是独立同分布的随机变量，第二是计算 $E[Y_1]$，应该是计算 $E[Y_i]$ 才对 —— 译者注。

② 原文在此好像没有说完——译者注。

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\alpha x)^2}{2\sigma^2}\right)$$

请计算条件均值 $E[Y|X]$，此处 α 是一个已知的常数。

4.31　我们希望利用函数 $p(x)$ 来估计 X 的概率密度函数，$p(x)$ 可使熵最大化。

$$H[X] \triangleq -\int_{-\infty}^{\infty} p(x) \ln p(x)\, \mathrm{d}x$$

已知 $E[X]=\mu$，$\mathrm{Var}[X]=\sigma^2$。请计算 X 概率密度函数的最大熵估计。

4.32　设 X: $N(0, \sigma^2)$，请证明下面的结论：

$$m_n \triangleq E[X^n] = 1\cdot 3\cdots(n-1)\sigma^n \qquad n\text{为偶数}$$

$$m_n = 0 \qquad n\text{为奇数}$$

4.33　已知 $\mu_X \triangleq E[X]$，$\mu_Y \triangleq E[Y]$，如果 $c_{11}=\sqrt{c_{20}c_{02}}$，请证明如下结论成立：

$$E\left[\left(\frac{c_{11}}{c_{20}}(X-\mu_X)-(Y-\mu_Y)\right)^2\right] = 0$$

根据上面这个结论，请证明当 $|\rho|=1$ 时，Y 是 X 的线性函数，即 $Y=\alpha X+\beta$，α 和 β 的取值与 X 和 Y 的各阶距有关。

4.34　对于例 4.3-4 中的最优线性预测器，最小均方差表达式如下：

$$\varepsilon_{\min}^2 = \sigma_Y^2(1-\rho^2)$$

请解释为什么 $|\rho|=1$ 时 $\varepsilon_{\min}^2=0$？

4.35　已知随机变量 X 的概率密度函数在 $0<x<2$ 区间为 $f_X(x)=1-(1/2)x$，其他区间 $f_X(x)=0$。请计算 X 的第 r 阶矩 m_r，r 是一个正整数。

4.36　设 $E[X_i]=\mu$，$\mathrm{Var}[X_i]=\sigma^2$。我们希望用样本均值来估计 μ。

$$\widehat{\mu}_N \triangleq \frac{1}{N}\sum_{i=1}^{N} X_i$$

假设 X_i 是相互独立的，请计算 $\hat{\mu}_N$ 的均值和方差，$i=1,\cdots,N$。

4.37　对于上一个问题，试问 N 要多大，下面的不等式才成立。

$$P[|\widehat{\mu}_N-\mu|>0.1\sigma] \leqslant 0.01$$

4.38　设 X 是区间 $\left(-\frac{1}{2}, \frac{1}{2}\right)$ 中均匀分布随机变量，请计算：(a) X 的矩量生成函数；(b) 利用式(4.5-5)计算 X 的均值。提示：$\sinh z \triangleq (\mathrm{e}^z-\mathrm{e}^{-z})/2$，取极限可计算均值。

4.39　设 X 是一个泊松分布的随机变量，请计算 (a) 矩量生成函数；(b) 利用式(4.5-5)来计算均值。

4.40　负二项分布的概率质量函数定义如下，其中 $Q-P=1$，$P>0$，$N\geqslant 1$。

$$P_X(k) \triangleq \binom{N+k-1}{N-1}\left(\frac{P}{Q}\right)^k\left(1-\frac{P}{Q}\right)^N \qquad (k=0,1,2,\cdots)$$

当不能确保各个事件是独立发生的（这是泊松分布的严格随机性要求），有时就用上面这个分布来代替泊松分布。请证明矩量生成函数是如下表达式：

$$M_X(t) = (Q-P\mathrm{e}^t)^{-N}$$

提示：可尝试计算或查表得到 $(Q-P\mathrm{e}^t)^{-N}$ 的展开公式，比如可参考 *Discrete Distributions*, N. L. Johnson and S. Kotz, John Wiley and Sons, 1969。

4.41　已知 X 服从伽马分布，概率密度函数如下，请计算 X 的矩量生成函数。

$$f_X(x;\alpha,\beta) = \begin{cases} (\alpha!\,\beta^{\alpha+1})^{-1} & x^\alpha \exp(-x/\beta),\ 0<x<\infty,\ \beta>0,\ \alpha\geqslant 0 \\ 0 & \text{其他} \end{cases}$$

4.42 如果 X 服从伽马分布，请计算 X 的均值和方差。

4.43 已知随机变量 X 服从指数分布，$f_X(x) = \lambda e^{-\lambda x}u(x)$，请计算关于 $P[X \geqslant a]$ 的切尔诺夫限。

4.44 设习题 4.40 中 $N = 1$。(a)计算 $P[X \geqslant k]$ 的切尔诺夫限；(b)将结论推广到任意的 N。

4.45 设 X 服从柯西分布，概率密度函数如下，试计算 X 的特征函数 $\Phi_X(\omega)$。

$$f_X(x) = \frac{\alpha}{\pi(\alpha^2 + x^2)}$$

4.46 设 X 服从柯西分布，概率密度函数为 $f_X(x) = (\pi(1 + (x - a)^2))^{-1}$，$-\infty < x < \infty$，请计算 $E[X]$。当你准备计算 σ_X^2 时会遇到什么问题？

4.47 请计算指数分布随机变量的特征函数，概率密度函数如下，其中均值 $\mu > 0$，$u(x)$ 表示单位阶跃函数。

$$f_X(x) = \frac{1}{\mu} e^{-x/\mu} u(x)$$

4.48 设 $Y = \frac{1}{N}\sum_{i=1}^{n} X_i$，其中 X_i 是独立同分布的柯西随机变量，概率密度函数为

$$f_{X_i}(x) = \frac{1}{\pi[1 + (x - \mu)^2]} \quad i = 1, \cdots, N$$

请证明 Y 的概率密度函数为

$$f_Y(x) = \frac{1}{\pi[1 + (x - \mu)^2]}$$

也就是说 Y 的概率密度函数与各个 X_i 的概率密度函数相同，并且与 N 无关(提示：可使用特征函数法)。

4.49 设 X 在区间 $(-a, a)$ 上服从均匀分布，Y 与 X 独立且在区间 $([n-2]a, na)$ 上服从均匀分布，$n = 1, 2, \cdots$。对于不同的 n，请计算 $Z = X + Y$ 的数学期望，并绘出 Z 的概率密度函数曲线。请问 n 在这里起了什么作用？

4.50 下面的递归表达式被称为一阶滑动平均，其中 X_n，Z_n 和 Z_{n-1} 都是随机变量，$n = \cdots, -1, 0, 1, \cdots$。

$$X_n = Z_n - aZ_{n-1}, \qquad |a| < 1$$

假设对于所有的 n，$E[Z_n] = 0$ 和 $E[Z_n^2] = \sigma^2$；对于所有的 $n \neq j$，$E[Z_n Z_j] = 0$。请计算 $R_n(k) \triangleq E[X_n X_{n-k}]$，$k = 0, \pm 1, \pm 2, \cdots$。

4.51 下面的递归表达式称为一阶自回归

$$X_n = bX_{n-1} + Z_n, \qquad |b| < 1$$

假设对于所有的 n，$E[Z_n] = 0$ 和 $E[Z_n^2] = \sigma^2$；对于所有的 $n \neq j$，$E[Z_n Z_j] = 0$。另外假设 $E[Z_n X_{n-j}] = 0$，$j = 1, 2, \cdots$。$E[X_n^2] \triangleq K$ 与 n 无关。请计算 $R_n(k) = E[X_n X_{n-k}]$，$k = \pm 1, \pm 2, \cdots$。

4.52 设 $Z \triangleq aX + bY$，$W \triangleq cX + dY$。请以 X 和 Y 的联合特征函数来表示 Z 和 W 的联合特征函数。

4.53 设 $f_{XY}(x, y) = 16\exp(-4[x + y])$，$x > 0$，$y > 0$。请计算 (X, Y) 的联合矩量生成函数和联合特征函数。

4.54 设 X 和 Y 是两个独立的泊松分布随机变量，概率质量函数如下：

$$P_X(k) = \frac{1}{k!} e^{-2} 2^k$$

$$P_Y(k) = \frac{1}{k!} e^{-3} 3^k$$

请利用矩量生成函数或特征函数来计算 $Z = X + Y$ 的概率质量函数。

4.55 假设你的公司生产面包烤箱，生产出来的烤箱具有凹痕或刮痕的概率为 $p = 0.05$，这些烤箱具有凹痕或刮痕的情况是相互独立的。在一个星期内你的公司可以生产 2000 个烤箱。请问一星期内生产出来的所有烤箱中，至少有 109 个烤箱有凹痕或刮痕的概率为多少？

4.56 网络中的信息长度 L(以字节为单位)可以用独立同分布的指数随机变量来建模，其概率分布函数如下：

$$P[L \leqslant l] \triangleq F_L(l) = \begin{cases} 1 - e^{-0.002l}, & l \geqslant 0 \\ 0, & l < 0 \end{cases}$$

(a) 如果一个文件要存储 400 个信息，请问这个文件长度的数学期望是多少(字节)？

(b) 对于 400 个随机选取的信息，存储长度超过 520 个字节的概率是多少？

4.57　对于 100 个独立同分布的随机变量，均值为 μ，方差为 σ^2，我们给出样本均值为 $\hat{\mu}_{100}$。请使用切比雪夫不等式来计算概率 $P[|\hat{\mu}_{100}-\mu|>\sigma/5]$ 的上限，请给出数值结果。

4.58　你的公司生产液晶显示器(LCD)面板。设一个面板存在一个坏像素点的概率为 $p=0.03$。假设各个面板存在坏像素点的事件是独立的。该公司一个星期可以生产 2000 个 LCD 面板。请使用中心极限定理，估计一个星期内出现 80 个坏面板的概率。

4.59　设 X_i 是独立同分布的伯努利分布随机变量，$i=1, \cdots, n$，$P_X(1)=p$，$P_X(0)=q=1-p$。假设事件 $\{1\}$ 代表成功，事件 $\{0\}$ 代表失败。

(a) 请证明 Z_n 是一个零均值，单位方差的高斯分布随机变量，其中 $W_i \triangleq (X_i-p)/\sqrt{pq}$，$n\gg 1$。

$$Z_n \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i$$

(b) 请计算在 n 次试验中 k 次成功的概率 P，其中 $n=2000$，$k=110, 130, 150$。(i) 请使用严格的伯努利分布表达式；(ii) 请用泊松分布来近似伯努利分布；(iii) 请用中心极限定理来近似计算。请编写 3 个 MATLAB 小程序来完成计算。请验证一下当正确的概率降低时，中心极限定理的误差会随之增加。

4.60　在第 1 章中，通过二项分布的近似解决了下面的问题。

假设在计算机程序中错误代码发生的情况为：一行代码中包含错误的概率为 $p=0.001$，因此没有错误的概率为 $q=0.999$。不同的代码行中错误出现的情况是独立的。在一个 1000 行代码的程序中，至少有 2 行代码出现错误的概率是多少？

可以用中心极限定理来近似解决这个问题吗？能或不能，都请给出理由。

4.61　假设我们拥有一个均匀随机数发生器(URNG)，并且用一个独立同分布的均匀分布随机数序列来进行建模，X_i 表示这个均匀随机数发生器的第 i 个输出值，$i=1, \cdots, n$。此外还假设 X_i 的概率密度函数为

$$f_{X_i}(X_i)=\frac{1}{a}\mathrm{rect}\left(\frac{x_i-a/2}{a}\right)$$

(a) 试证明 $E[Z_n]=na/2$，其中 $Z_n=X_1+\cdots+X_n$。

(b) 请证明 $\mathrm{Var}[Z_n]=na^2/12$。

(c) 请编写一个 MATLAB 程序，绘出 $f_{Z_n}(z)$ 的曲线，其中 n 取 2, 3, 10, 20。

(d) 请编写一个 MATLAB 程序，绘出高斯概率密度函数 $N\left(\frac{na}{2}, \frac{na^2}{12}\right)$ 的曲线，其中 n 取 2, 3, 10, 20，并对所得的结果分别与 $f_{Z_n}(z)$ 进行比较。

(e) 对于不同的 n，请计算 $P[\mu_n-k\sigma_n\leqslant Z_n\leqslant\mu_n+k\sigma_n]$，其中 $\mu_n=na/2$，$\sigma_n^2=na^2/12$，k 的取值可以选择 0.1, 0.5, 1, 2, 3。请分别用 $f_{Z_n}(z)$ 和 $N(\frac{na}{2}, \frac{na^2}{12})$ 进行计算。a 可以选取任意合理的值，如可设 $a=1$。

4.62　设 $f_X(x)$ 是一个连续实随机变量的概率密度函数，如果 $f_X(x)=f_X(-x)$，试证明 $E[X]=0$。

4.63　请计算 Chi 平方随机变量 $W_n=\sum_{i=1}^{n}X_i^2$ 的方差，其中 X_i 是相互独立的，服从分布 $N(0,1)$，$i=1,2,\cdots,n$（提示：可使用习题 4.32 的结论）。

4.64　设 X 是一个高斯分布随机变量，X: $N(\mu, \sigma^2)$。请证明 $E\{(X-\mu)^{2k+1}\}=0$，$E[(X-\mu)^{2k}]=[(2k)!/2^k k!]\sigma^{2k}$。

4.65　(a) 请编写一个 MATLAB 程序(.m 文件)，绘出 Chi 平方分布随机变量 Z_n 的概率密度函数曲线，其中 $n=30, 40, 50$；(b) 改进你所编写的程序，计算 $P[\mu-\sigma\leqslant Z_n\leqslant\mu+\sigma]$。将得到的结果与高斯近似 $P[\mu-\sigma\leqslant X\leqslant\mu+\sigma]$ 相比较，其中 X: $N(n, 2n)$。

4.66 设 X_i 是 4 个零均值的高斯分布随机变量，$i=1,\cdots,4$。请利用联合特征函数证明下面结论成立。

$$E[X_1X_2X_3X_4]=E[X_1X_2]E[X_3X_4]+E[X_1X_3]E[X_2X_4]+E[X_2X_3]E[X_1X_4]$$

4.67 请计算 n 个自由度的 Chi 平方分布的矩量生成函数和特征函数。

4.68 设 $E[X_i]=\mu$，$\text{Var}[X_i]=\sigma^2$，我们希望用样本均值 $\hat{\mu}$ 来估计 μ

$$\hat{\mu}\triangleq\frac{1}{N}\sum_{i=1}^{N}X_i$$

请计算 $\hat{\mu}$ 的均值二阶矩，假设 X_i 是相互独立的，$i=1,\cdots,N$。

4.69 习题 4.62 的逆结论成立吗？也就是如果 $E[X]=0$，$f_X(x)=f_X(-x)$ 是否成立？

4.70 设 $X:U(a,b)$，$0<a<b$，$r\triangleq b/a$。当 $r\to\infty$ 时，请证明 μ/σ（均值-标准偏差比值）趋向于 1.73。

4.71 假设 X_i 是独立同分布的高斯随机变量，请证明 $W_n\triangleq\sum_{i=1}^{n}\left[(X_i-\frac{1}{n}\sum_{j=1}^{n}X_j)/\sigma\right]^2$ 服从 $n-1$ 个自由度的 Chi 平方分布。

4.72 (条件期望)设 $Y=X+N$，其中 X 和 N 是独立的泊松分布随机变量，均值分别为 20 和 5。

(a)请计算 Y 在 X 给定时的条件概率质量函数。

(b)请计算条件均值 $E[Y|X=x]$。

4.73 如果 $f_X(x)=f_X(-x)$，请证明不等式 $\sigma_XP[|X|\geqslant\sigma_X]\leqslant E[|X|]\leqslant\sigma_X$ 成立。

4.74 假设 μ_X，μ_Y，σ_X^2，σ_Y^2 和 ρ 已知，对于随机变量 X 和 Y，我们用线性方程来估计 Y

$$\hat{Y}=\alpha X+\beta$$

定义估计误差如下：

$$\varepsilon\triangleq\hat{Y}-Y$$

(a)请计算估计误差 ε 与数据 X 之间的协方差，即

$$\text{Cov}[\varepsilon,X]=E[\varepsilon X]-E[\varepsilon]E[X]$$

请用 α，β 以及本题给出的其他参数来进行计算。

(b)请给出 α 和 β 的最优值，然后再次计算 $\text{Cov}[\varepsilon,X]$。

4.75 在习题 4.74 中，利用随机变量 X 来对随机变量 Y 进行线性估计

$$\hat{Y}=\alpha X+\beta$$

最优取值的 α 和 β 会使得 $\text{Cov}[\varepsilon,X]=0$。下面的这个条件就是所谓的正交条件

$$\text{Cov}[\varepsilon,X]=0$$

请证明对于一般的线性估计问题，只要满足正交条件，所得线性估计的均方误差是最小的。也就是说，此时估计误差 ε 与观测数据 X 是正交的。

在这里，我们考虑一个更一般的问题。利用两个随机变量 X_1 和 X_2 来线性估计 Y，也就是

$$\hat{Y}=\alpha_1X_1+\alpha_2X_2+\beta$$

采用下面的正交条件来计算 α_1 和 α_2 的取值。

$$\text{Cov}[\varepsilon,X_1]=0\quad\text{和}\quad\text{Cov}[\varepsilon,X_2]=0$$

为使问题更简单一些，假设三个均值都是零，此时 $\beta=0$，所以线性估计器可简化为

$$\hat{Y}=\alpha_1X_1+\alpha_2X_2$$

同前面类似，估计误差为 $\varepsilon=Y-\hat{Y}$。因为均值是零，所以 $\text{Cov}[\varepsilon,X_1]=E[\varepsilon X_1]$，$\text{Cov}[\varepsilon,X_2]=E[\varepsilon X_2]$。请使用下面的参数设置：

$$\sigma_1^2=1,\sigma_2^2=4,\sigma_Y^2=4$$
$$\rho_1=0.5,\rho_2=0.7,\rho_{12}=0.5$$

其中 $\rho_1=E[X_1Y]/\sigma_1\sigma_Y$，$\rho_2=E[X_2Y]/\sigma_2\sigma_Y$，$\rho_{12}=E[X_1X_2]/\sigma_1\sigma_2$。

（a）请根据上面给出的参数写出两个线性方程，根据下面的正交条件，可以用这两个方程来求解 α_1 和 α_2。

$$E[\varepsilon X_1]=0 \quad 和 \quad E[\varepsilon X_2]=0$$

（b）求解这两个线性方程，解出 α_1 和 α_2。

参考文献

4-1 W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edition. New York: John Wiley, 1971.

4-2 W. B. Davenport, Jr., *Probability and Random Processes: An Introduction for Applied Scientists and Engineers*, New York: McGraw-Hill, p. 99.

4-3 A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd edition, New York: McGraw-Hill, 1991.

4-4 B. Saleh, *Photoelectron Statistics*. New York: Springer-Verlag, 1978, Chapter 5.

4-5 R. G. Gallagher, *Information Theory and Reliable Communications*. New York: John Wiley, 1968.

4-6 P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Part 1. New York: McGraw-Hill, 1953, p. 279.

4-7 G. A. Korn and T. S. Korn, *Mathematical Handbook for Scientists and Engineers*. New York: McGraw-Hill, 1961.

4-8 W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd edition. New York: John Wiley, 1957.

补充文献

Cooper, G. R. and C. D. McGillem, *Probabilistic Methods of Signal and System Analysis*, 3rd edition. New York: Holt, Rinehart and Winston, 1999.

Peebles, P. Z. Jr., *Probability, Random Variables, and Random Signal Principles*, 4th edition. New York: McGraw-Hill, 2001.

Garcia, L.-G., *Probability and Random Processes for Electrical Engineering*, 2nd edition. Reading, MA: Addison-Wesley, 1994.

Helstrom, C. W., *Probability and Stochastic Processes for Engineers*, 2nd edition. New York: Macmillan, 1991.

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd edition. New York: McGraw-Hill, 1991.

Scheaffer, R. L., *Introduction to Probability and Its Applications*. Belmont, CA: Duxbury, 1990.

Viniotis, Y., *Probability and Random Processes for Electrical Engineers*. New York: McGraw-Hill, 1998.

Yates, R. D. and D. J. Goodman, *Probability and Stochastic Processes*. New York: Wiley, 1999.

第5章 随机向量

5.1 联合分布和联合密度

在许多涉及随机现象的问题中，我们观测到它们往往具有向量的特征。对此用三个例题来进行讲解。

例 5.1-1 （地震鉴别）一个地震监视台接收到一个地震波 $X(t)$，在时刻 t_1，t_2，…，t_n 进行采样，可得到一个向量 $\boldsymbol{X}=[X_1, \cdots, X_n]^{\mathrm{T}}$，其中 $X_i \triangleq X(t_i)$，T 表示转置操作①。因为政治和军事上的原因，在过去有段时间，确定这个波是来自地震还是地下爆炸非常重要。假设有套专家电脑系统，这套系统存储了很多的数据，既包括地震，也包括地下爆炸。将向量 $\boldsymbol{X}$ 与存储的数据相比较，请问 $X(t)$ 可以被准确鉴定的概率是多少？

例 5.1-2 （健康向量）一个地区的卫生部门为了评估小学儿童的健康状况，需要测量每个儿童的一些数据，包括身高，体重，血压，红血球数目，白血球数目，肺活量，心率，血铅水平，视觉灵敏度等。得到的向量 $\boldsymbol{X}$ 可作为每个儿童健康状况的一个总结。请问随机选取的儿童中健康儿童的概率是多少？

例 5.1-3 （疾病监测）一个配备了数字扫描仪的电脑系统，可以通过 X 射线来检测黑肺病。通过统计 6 个肺部区域(即每片肺叶 3 个区域)中射线无法通过的数目，估计每个区域中阴影区域的平均面积。得到的结果是一个 12 维的向量 $\boldsymbol{X}$，用这个向量来进行诊断。请问计算机最好的诊断决策应该是什么？

前面的三个例题在工程和科学问题中非常具有代表性，为了某种目的需要把一些随机变量进行编组。用向量的方法来学习这些编组的随机变量非常方便。因此，我们把这些编组的随机变量看成一个整体，就称为随机向量。如前面章节讲到的，字母表后面的大写字母表示随机变量，黑斜体的大写字母表示随机向量和矩阵，小写的黑斜体字母表示确定性向量，即一个随机向量的具体取值。

考虑一个样本描述空间 Ω 和其中的点 ζ，以及从 Ω 映射到实数直线随机向量(RV)中 n 个随机变量组成的集合 $X_1, X_2, \cdots, X_n$。对于每个 $\zeta \in \Omega$，可以生成一个由 n 个元素组成的向量 $\boldsymbol{X}(\zeta) \triangleq (X_1(\zeta), X_2(\zeta), \cdots, X_n(\zeta)) \in R^n$，那么 $\boldsymbol{X} \triangleq (X_1, X_2, \cdots, X_n)$ 就是一个 n 维实随机向量。该定义也可拓展到复随机向量。设 $\boldsymbol{X}$ 是定义在样本空间 Ω 上的 n 维随机向量，概率分布函数为 $F_{\boldsymbol{X}}(\boldsymbol{x})$。由定义可知②

$$F_{\boldsymbol{X}}(\boldsymbol{x}) \triangleq P[X_1 \leqslant x_1, \cdots, X_n \leqslant x_n] \tag{5.1-1}$$

---

① 除非特别说明，本书中所有的向量都是列向量。

② 在此提醒读者：事件$(\boldsymbol{X} \leqslant \boldsymbol{x}) \triangleq \{X_1 \leqslant x_1, \cdots, X_n \leqslant x_n\}$就是 n 个事件$\{X_i \leqslant x_i\}$的交集，$i=1, \cdots, n$。如果其中任何一个子事件是不可能事件，比如$\{X_i \leqslant -\infty\}$，那么整个事件就是一个不可能事件，此时 $F_{\boldsymbol{X}}(-\infty)=0$ 仍然成立。

通过定义 $\{\boldsymbol{X} \leqslant \boldsymbol{x}\} \triangleq \{X_1 \leqslant x_1, \cdots, X_n \leqslant x_n\}$，可把式(5.1-1)简写为

$$F_{\boldsymbol{X}}(\boldsymbol{x}) \triangleq P[\boldsymbol{X} \leqslant \boldsymbol{x}] \tag{5.1-2}$$

我们认为事件 $\{\boldsymbol{X} \leqslant \infty\}$ 和 $\{\boldsymbol{X} \leqslant -\infty\}$ 分别是必然事件和不可能事件，因此有

$$F_{\boldsymbol{X}}(\infty) = 1 \tag{5.1-3a}$$

$$F_{\boldsymbol{X}}(-\infty) = 0 \tag{5.1-3b}$$

如果 $F_{\boldsymbol{X}}(\boldsymbol{x})$ 的 n 维混合偏导存在，就可以定义概率密度函数如下：

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \triangleq \frac{\partial^n F_{\boldsymbol{X}}(\boldsymbol{x})}{\partial x_1 \cdots \partial x_n} \tag{5.1-4}$$

读者可能会发现这些定义与第2章中标量定义是完全类似的，因此可以通过如下定义得到式(5.1-4)。

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \triangleq \lim_{\substack{\Delta x_1 \to 0 \\ \vdots \\ \Delta x_n \to 0}} \frac{P[x_1 < X_1 \leqslant x_1 + \Delta x_1, \cdots, x_n < X_n \leqslant x_n + \Delta x_n]}{\Delta x_1 \cdots \Delta x_n} \tag{5.1-5}$$

举个例子，当 $n=2$ 时

$$\begin{aligned} &P[x_1 < X_1 \leqslant x_1 + \Delta x_1, x_2 < X_2 \leqslant x_2 + \Delta x_2] \\ &\quad = F_{\boldsymbol{X}}(x_1 + \Delta x_1, x_2 + \Delta x_2) - F_{\boldsymbol{X}}(x_1, x_2 + \Delta x_2) - F_{\boldsymbol{X}}(x_1 + \Delta x_1, x_2) + F_{\boldsymbol{X}}(x_1, x_2) \end{aligned}$$

因此可得

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) = \lim_{\substack{\Delta x_1 \to 0 \\ \Delta x_2 \to 0}} \frac{1}{\Delta x_1 \Delta x_2} &[F_{\boldsymbol{X}}(x_1 + \Delta x_1, x_2 + \Delta x_2) - F_{\boldsymbol{X}}(x_1 + \Delta x_1, x_2) \\ &- F_{\boldsymbol{X}}(x_1, x_2 + \Delta x_2) + F_{\boldsymbol{X}}(x_1, x_2)] \end{aligned}$$

根据定义可知，上式是函数的二阶混合偏导，因此有

$$f_{\boldsymbol{X}}(x_1, x_2) = \frac{\partial^2 F_{\boldsymbol{X}}(x_1, x_2)}{\partial x_1 \partial x_2}$$

当增量很小的时候，从式(5.1-5)可得到如下有用的等式

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \Delta x_1 \cdots \Delta x_n \approx P[x_1 < X_1 \leqslant x_1 + \Delta x_1, \cdots, x_n < X_n \leqslant x_n + \Delta x_n] \tag{5.1-6}$$

将式(5.1-4)进行积分可以得到概率分布函数如下：

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\boldsymbol{X}}(\boldsymbol{x}')\, \mathrm{d}x_1' \cdots \mathrm{d}x_n'$$

将其写成紧凑的表达式为

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{\boldsymbol{x}} f_{\boldsymbol{X}}(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}'$$

更一般地，对于任意一个事件 $B \subset \boldsymbol{R}^N$（$\boldsymbol{R}^N$ 是一个 N 维欧式空间），其包含可数集和平行六面体交集，有如下结论成立：

$$P[B] = \int_{\boldsymbol{x} \in B} f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{5.1-7}$$

［与式(2.5-3)相比较］式(5.1-7)参数的合理性与一维的情况非常类似（参见2.5节）。Davenport[5-1, p.149]讨论过式(5.1-7)在 $n=2$ 时的参数合理性。对于 $n>2$ 的情况，可以用数学归纳法证明。

$\boldsymbol{X}$ 关于事件 B 的概率分布函数定义如下：

$$F_{\boldsymbol{X}|B}(\boldsymbol{x}|B) \triangleq P[\boldsymbol{X} \leqslant \boldsymbol{x}|B] = \frac{P[\boldsymbol{X} \leqslant \boldsymbol{x}, B]}{P[B]} \quad (P[B] \neq 0)$$

上面的这些结果，以及接下来的结果与一维的情况很类似。考虑 n 的不相交的穷举事件 $\{B_i, i=1, \cdots, n\}$，且 $P[B_i]>0$。此时有 $\cup_{i=1}^{n} B_i = \Omega$, $B_iB_j = \phi$, $i \neq j$。根据全概率定理 1.6-1 可知

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{i=1}^{n} F_{\boldsymbol{X}|B_i}(\boldsymbol{x}|B_i)P[B_i] \tag{5.1-8}$$

等式左边的这个条件概率分布函数有的时候也称为混合分布函数。$\boldsymbol{X}$ 关于事件 B 的条件概率密度函数是 $F_{\boldsymbol{X}|B}(\boldsymbol{x}|B)$ 的 n 阶混合偏导(前提是这个偏导存在)，表达式如下：

$$f_{\boldsymbol{X}|B}(\boldsymbol{x}|B) \triangleq \frac{\partial^n F_{\boldsymbol{X}|B}(\boldsymbol{x}|B)}{\partial x_1 \cdots \partial x_n} \tag{5.1-9}$$

根据式(5.1-8)和式(5.1-9)可得

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{i=1}^{n} f_{\boldsymbol{X}|B}(\boldsymbol{x}|B_i)P[B_i] \tag{5.1-10}$$

因为 $f_{\boldsymbol{X}}(\boldsymbol{x})$ 是混合的，即它是条件概率密度函数的线性组合，有的时候也被称为混合概率密度函数①。

两个随机向量 $\boldsymbol{X}=[X_1, \cdots, X_n]^{\mathrm{T}}$ 和 $\boldsymbol{Y}=[Y_1, \cdots, Y_m]^{\mathrm{T}}$ 的联合概率分布函数为

$$F_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y}) = P[\boldsymbol{X} \leqslant \boldsymbol{x}, \boldsymbol{Y} \leqslant \boldsymbol{y}] \tag{5.1-11}$$

如果 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的联合概率密度函数存在，那么有如下定义：

$$f_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y}) = \frac{\partial^{(n+m)} F_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y})}{\partial x_1 \cdots \partial x_n \partial y_1 \cdots \partial y_m} \tag{5.1-12}$$

$\boldsymbol{X}$ 的边缘概率密度函数 $f_{\boldsymbol{X}}(\boldsymbol{x})$ 可通过对 $f_{\boldsymbol{XY}}(\boldsymbol{x}, \boldsymbol{y})$ 积分得到

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}y_1 \cdots \mathrm{d}y_m$$

类似地，简化向量 $\boldsymbol{X}' \triangleq [X_1, \cdots, X_{n-1}]^{\mathrm{T}}$ 的边缘概率密度函数可由 $\boldsymbol{X}$ 的概率密度函数得到

$$f_{\boldsymbol{X}'}(\boldsymbol{x}') \triangleq \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}x_n \qquad \text{其中 } \boldsymbol{x}' \triangleq [x_1, \cdots, x_{n-1}]^{\mathrm{T}} \tag{5.1-13}$$

很显然，式(5.1-13)可以拓展到所有的边缘概率密度函数，只需要对相应的变量进行积分即可。

例 5.1-4 (随机粒子)设 $\boldsymbol{X}=[X_1, X_2, X_3]^{\mathrm{T}}$ 表示一个粒子的位置，该粒子在一个球心在原点，半径为 a 的球体内。假设在观测的时刻，粒子等概率的出现在球体任何一个位置

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \begin{cases} \dfrac{3}{4\pi a^3}, & \sqrt{x_1^2+x_2^2+x_3^2} < a \\ 0, & \text{其他} \end{cases}$$

请计算该粒子出现在一个半径为 $2a/3$ 的子球体内的概率为多少。

---

① 这在统计模式识别中经常用到。

解 假设 E 代表粒子出现在这个子球体内的这个事件(为简单起见，假设子球体的球心也在原点)，设

$$\mathscr{R} \triangleq \{x_1, x_2, x_3 : \sqrt{x_1^2 + x_2^2 + x_3^2} < 2a/3\}$$

对于下面等式的计算，用球坐标系表示是最方便的

$$P[E] = \iiint f_{\boldsymbol{X}}(x_1, x_2, x_3)\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3$$

也就是

$$P[E] = \frac{3}{4\pi a^3} \int_{r=0}^{2a/3} \int_{\phi=0}^{\pi} \int_{\theta=0}^{2\pi} r^2 \sin\phi\,\mathrm{d}r\,\mathrm{d}\phi\,\mathrm{d}\theta$$

在这个简单的例题里面，最终结果是体积之比，也就是 $(2a/3)^3 \div a^3 = 8/27 \approx 0.3$。

5.2 随机变量的多重变换

本节是对3.4节内容的一个直接拓展。设 $\boldsymbol{X}$ 是定义在样本空间 Ω 上的一个 n 维随机向量。对于下面的 n 个实函数

$$\begin{aligned} y_1 &= g_1(x_1, x_2, \cdots, x_n) \\ y_2 &= g_2(x_1, x_2, \cdots, x_n) \\ &\vdots \\ y_n &= g_n(x_1, x_2, \cdots, x_n) \end{aligned} \tag{5.2-1}$$

其中 g_i，$i = 1, \cdots, n$ 是函数无关的，表示不存在恒等于0的函数 $H(y_1, y_2, \cdots, y_n)$。例如，下面三个线性方程就不是函数无关的：

$$\begin{aligned} y_1 &= x_1 - 2x_2 + x_3 \\ y_2 &= 3x_1 + 2x_2 + 2x_3 \\ y_3 &= 5x_1 - 2x_2 + 4x_3 \end{aligned} \tag{5.2-2}$$

因为对于 x_1, x_2, x_3 的所有取值，函数 $H(y_1, y_2, \cdots, y_n) = 2y_1 + y_2 - y_3 = 0$。根据下面的等式，可以建立一个 n 维的随机向量 $\boldsymbol{Y} \triangleq [Y_1, Y_2, Y_n]$

$$\begin{aligned} Y_1 &= g_1(X_1, X_2, \cdots, X_n) \\ Y_2 &= g_2(X_1, X_2, \cdots, X_n) \\ &\vdots \\ Y_n &= g_n(X_1, X_2, \cdots, X_n) \end{aligned} \tag{5.2-3}$$

利用这种方法，生成了 n 个函数，每个函数由 n 个随机变量组成。为了行文简洁，我们设 $\boldsymbol{x} \triangleq [x_1, x_2, \cdots, x_n]$，$\boldsymbol{y} \triangleq [y_1, y_2, \cdots, y_n]$。请问：假设已知联合概率密度函数 $f_{\boldsymbol{X}}(\boldsymbol{x})$，如何计算 Y_i 的联合概率密度函数 $f_{\boldsymbol{Y}}(\boldsymbol{y})$，$i = 1, \cdots, m$？如果随机变量 $Y_i(i = 1, \cdots, m)$ 的数目比 $X_i(i = 1, \cdots, n)$ 少，即 $m < n$，此时可以仿照例3.5-4通过辅助方程来增加 Y_i。

假设对于方程组(5.2-1)，可以求得 x_i 的唯一解如下（$i = 1, \cdots, n$）：

$$\begin{aligned} x_1 &= \phi_1(y_1, y_2, \cdots, y_n) \\ x_2 &= \phi_2(y_1, y_2, \cdots, y_n) \\ &\vdots \\ x_n &= \phi_n(y_1, y_2, \cdots, y_n) \end{aligned} \tag{5.2-4}$$

对于无穷小事件 $A \triangleq \{\zeta: y_i < Y_i \leqslant y_i + \mathrm{d}y_i,\ i = 1, \cdots, n\}$, Y_i 只能在无穷小的长方体 $\mathscr{P}_y$ 内取值。按照式(3.4-5)至式(3.4-8)的求解过程，可以得到

$$P[A] = \int_{\mathscr{P}_y} f_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}y = f_{\boldsymbol{Y}}(\boldsymbol{y})\,V_y = \int_{\mathscr{P}_x} f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}x = f_{\boldsymbol{X}}(\boldsymbol{x})\,V_x \tag{5.2-5}$$

其中 $\mathscr{P}_x$ 是一个无穷小的平行六面体(不一定要求是长方体)，V_y 是 $\mathscr{P}_y$ 的体积，V_x 是 $\mathscr{P}_x$ 的体积。根据式(5.2-5)可得

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x})\frac{V_x}{V_y} \tag{5.2-6}$$

根据附录 C 可知，无穷小体积之比表示因子 $\tilde{J}$ 的大小，具体为

$$\tilde{J} = \begin{vmatrix} \dfrac{\partial\phi_1}{\partial y_1} & \cdots & \dfrac{\partial\phi_1}{\partial y_n} \\ \vdots & & \vdots \\ \dfrac{\partial\phi_n}{\partial y_1} & \cdots & \dfrac{\partial\phi_n}{\partial y_n} \end{vmatrix} \tag{5.2-7}$$

$$= \begin{vmatrix} \dfrac{\partial g_1}{\partial x_1} & \cdots & \dfrac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial g_n}{\partial x_1} & \cdots & \dfrac{\partial g_n}{\partial x_n} \end{vmatrix}^{-1} = J^{-1} \tag{5.2-8}$$

因此可得

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x})|\tilde{J}| = f_{\boldsymbol{X}}(\boldsymbol{x})/|J| \tag{5.2-9}$$

一般而言，$\boldsymbol{y}$ 系统中的一个无穷小的长方体映射到 $\boldsymbol{x}$ 系统中 r 个不相交的无穷小的平行六面体。前面提到的事件 A 就是事件 E_i, $i = 1, \cdots, r$①的并集，其中 $E_i = \{\boldsymbol{X} \in \mathscr{P}_x^{(i)}\}$, $\mathscr{P}_x^{(i)}$ 的体积为 $V_x^{(i)}$，是 $\boldsymbol{x}$ 系统中 r 个平行六面体之一。因为这些事件(区域)是不相交的，将基本概率 $P[E_i]$ 相加，就可得到本节的最终结果如下：

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{i=1}^{r} f_{\boldsymbol{X}}(\boldsymbol{x}^{(i)})|\tilde{J}_i| \tag{5.2-10}$$

$$= \sum_{i=1}^{r} f_{\boldsymbol{X}}(\boldsymbol{x}^{(i)})/|J_i| \tag{5.2-11}$$

在式(5.2-10)和式(5.2-11)中，$|\tilde{J}_i| \triangleq V_x^{(i)}/V_y$, $|\tilde{J}_i| = |J_i|^{-1}$。

例 5.2-1 (向量变换)已知向量 $\boldsymbol{x}$ 的三个标量变换如下：

$$g_1(\boldsymbol{x}) = x_1^2 - x_2^2$$
$$g_2(\boldsymbol{x}) = x_1^2 + x_2^2$$
$$g_3(\boldsymbol{x}) = x_3$$

对于如下的系统，有 4 个解(根)存在

$$y_1 = x_1^2 - x_2^2$$

① 原文有误，在这里是 $i = l, \cdots, r$——译者注。

$$y_2 = x_1^2 + x_2^2$$
$$y_3 = x_3$$

这4个根分别是

$$\begin{aligned}
&x_1^{(1)} = ((y_1+y_2)/2)^{1/2} && x_1^{(2)} = ((y_1+y_2)/2)^{1/2} \\
&x_2^{(1)} = ((y_2-y_1)/2)^{1/2} && x_2^{(2)} = -((y_2-y_1)/2)^{1/2} \\
&x_3^{(1)} = y_3 && x_3^{(2)} = y_3 \\
&x_1^{(3)} = -((y_1+y_2)/2)^{1/2} && x_1^{(4)} = -((y_1+y_2)/2)^{1/2} \\
&x_2^{(3)} = ((y_2-y_1)/2)^{1/2} && x_2^{(4)} = -((y_2-y_1)/2)^{1/2} \\
&x_3^{(3)} = y_3 && x_3^{(4)} = y_3
\end{aligned} \tag{5.2-12}$$

为了使根都是实数，$y_2 \geqslant 0$，$y_1+y_2 \geqslant 0$，$y_2-y_1 \geqslant 0$，因此有$y_2 \geqslant |y_1|$。在这道例题里面，三维空间$\boldsymbol{y}$中的一个无穷小的立方体，映射到三维空间$\boldsymbol{x}$中的4个不相交的无穷小平行六面体。

例5.2-2　(更多的向量变换)对于例5.2-1中的向量变换，请根据$f_X(x)$的表达式计算$f_{\boldsymbol{Y}}(\boldsymbol{y})$

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-3/2}\exp\left[-\frac{1}{2}(x_1^2+x_2^2+x_3^2)\right]$$

其中$\boldsymbol{X}$是一个三维的标准正态分布随机变量。

解　必须首先计算这4个根各自的雅可比值$|J|$

$$J = \begin{vmatrix} 2x_1 & -2x_2 & 0 \\ 2x_1 & +2x_2 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 8x_1x_2$$

例如，第一个根的雅可比值可计算为

$$J_1 = 4(y_2^2 - y_1^2)^{1/2}$$

直接计算可知$|J_1| = |J_2| = |J_3| = |J_4|$。最后，将式(5.2-12)中的4个根标记为$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，$\boldsymbol{x}_3$，$\boldsymbol{x}_4$，可得

$$\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \frac{1}{4(y_2^2-y_1^2)^{1/2}}\sum_{i=1}^{4} f_{\boldsymbol{X}}(\boldsymbol{x}_i) \\
&= \frac{(2\pi)^{-3/2}}{(y_2^2-y_1^2)^{1/2}}\exp\left[-\frac{1}{2}(y_2+y_3^2)\right] \times u(y_2)u(y_2-|y_1|)\text{①}
\end{aligned}$$

一个随机向量完全可以由其分布函数或密度函数描述其特性，但除了几个特例外，用密度函数是很困难的。到目前为止，两个最重要的特例是：(1) $F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{X_1}(x_1)\cdots F_{X_n}(x_n)$，即$\boldsymbol{X}$的$n$个分量都是独立的；(2)如果$\boldsymbol{X}$服从多维高斯分布的规律。第1种情况是很好解决的，可以将其看成标量情况的一种直接拓展。第2种情况的解决将在5.6节中讨论。但是，如果我们所遇到的既不是第1种也不是第2种情况时，该如何应对呢？估计非独立变量的多维分布是不现实的，即使能够计算也是非常复杂的，没有任何实用价值。因此对于随机向量，一般倾向使用基于各阶距的特性，这些特性虽然不太完美但更具有可计算性。对于绝大多数工程应

---

① 证明这个概率密度函数的积分为1有些复杂。

用，最重要的矩一般就是期望向量(一阶矩)和协方差矩阵(二阶矩)。它们的特性和应用将在后面的章节讨论。接下来，将学习具有有序分量的随机向量。

5.3　有序随机变量

在3.4节(例3.4-3和例3.4-5)中，介绍了两个有序随机变量的符号。在此，生成 n 个随机变量，并以此得到它们的一些重要特性。有序随机变量是非常重要的，因为在无法了解随机变量分布特性的情况下，有序变换的统计知识可以告诉我们很多有用的信息，比如中位数，取值范围，以及其他一些与分布参数紧密联系的信息。对于 n 个独立同分布的连续随机变量，每个变量的概率密度函数为 $f_X(x)$，$-\infty < x < \infty$。这 n 个随机变量的联合概率密度函数为 $f_{X_1\cdots X_n}(x_1, \cdots, x_n) = f_X(x_1)\cdots f_X(x_n)$，联合边缘概率密度函数可由积分得到，比如 X_1 和 X_n 的联合边缘概率密度函数可对变量 $x_2, \cdots, x_{n-1}$ 积分得到。现在将这 n 个随机变量按照升序排列，即如果 $X_k = \min(X_1, \cdots, X_n)$，那么 $Y_1 = X_k$，此时 Y_2 就是接下来的集合 $\{X_i, i = 1, \cdots, n, i \neq k\}$ 中最小的，Y_3 又是再接下来的集合中最小的，以此类推，最后 $Y_n = \max(X_1, \cdots, X_n)$。此时就完成了一个有序变换，因为假设 X_i 是连续的随机变量，因此可以写出一个依概率1严格成立不等式 $Y_1 < Y_2 < \cdots < Y_{n-1} < Y_n$。我们希望计算 $\{Y_i, i = 1, \cdots, n\}$ 的联合概率密度函数。乍看上去，可能会错误地认为：因为集合 $S_1 = \{X_i, i = 1, \cdots, n\}$ 和 $S_2 = \{Y_i, i = 1, \cdots, n\}$ 中的元素是完全相同的，所以当 $\{y_i: -\infty < y_i < \infty, i = 1, \cdots, n\}$ 时有下面的结论成立：

$$\begin{aligned} f_{Y_1\cdots Y_n}(y_1,\cdots,y_n) &= f_{X_1\cdots X_n}(y_1,\cdots,y_n) \\ &= f_{X_1}(y_1)\cdots f_{X_n}(y_n) \end{aligned}$$

但是，这个结果忽略了一个重要的事实，那就是 $\{Y_i, i = 1, \cdots, n\}$ 不是独立的随机变量。打个比方，如果观测到了 X_1，那么能从 X_1 中了解到 X_2 的什么特性呢？你什么都不会了解到的。但是，如果已知了 Y_1，马上就会知道 $Y_2 > Y_1$，而且同样也会知道 $Y_1 > Y_2$ 的概率为0，因为在 $y_1 > y_2$ 的区间概率为0。根据这个情况，我们可能希望将 $\{Y_i\}$ 的联合概率密度函数修改为下面的形式：

$$\begin{aligned} f_{Y_1\cdots Y_n}(y_1,\cdots,y_n) &= f_{X_1}(y_1)\cdots f_{X_n}(y_n), \qquad y_1 < y_2 < \cdots y_n \\ &= 0, \text{ 其他} \end{aligned}$$

不过，我们又遇到了新的问题：修改后的联合概率密度函数其包围的体积不是1。实际上当 n 比较大的时候，体积是小于1的。为了得到 $\{Y_i\}$ 联合概率密度函数的正确表达式，我们将用到5.2节中的结论，根据这个结论可以计算一个随机变量集合的概率密度函数，这些变量的概率密度函数是已知的，且存在函数关系。

首先把一个 n 维空间 $\{-\infty < x_1, x_2, \cdots, x_n < \infty\}$ 划分为 $n!$ 个非重叠的不同区域，用 $\mathscr{R}_i = \{x_{i(1)} < x_{i(2)} < \cdots x_{i(j)} < \cdots < x_{i(n)}\}$ 表示，其中 $1 \leqslant i \leqslant n!$，$1 \leqslant j \leqslant n$，$x_{i(j)} \in \{x_1, x_2, \cdots, x_n\}$。而且当 $j < k$ 时 $x_{i(j)} < x_{i(k)}$。每个区域中元素的排序都不相同。比如，对于一个三维空间 (x_1, x_2, x_3)，它的非重叠，不相同可划分为

$$\begin{aligned} \mathscr{R}_1 &= \{x_1 < x_2 < x_3\} \\ \mathscr{R}_2 &= \{x_1 < x_3 < x_2\} \\ \mathscr{R}_3 &= \{x_2 < x_1 < x_3\} \\ \mathscr{R}_4 &= \{x_2 < x_3 < x_1\} \\ \mathscr{R}_5 &= \{x_3 < x_1 < x_2\} \\ \mathscr{R}_6 &= \{x_3 < x_2 < x_1\} \end{aligned}$$

对于$n!$个区域中的每个区域，我们定义$y_1 \triangleq x_{j(1)} < y_2 \triangleq x_{j(2)} < \cdots y_n \triangleq x_{j(n)}$，$j=1, \cdots, n!$。以三维空间$(x_1, x_2, x_3)$为例

$$\begin{aligned}
&\mathscr{R}_1: y_1 = x_1; y_2 = x_2; y_3 = x_3 \\
&\mathscr{R}_2: y_1 = x_1; y_2 = x_3; y_3 = x_3 \\
&\mathscr{R}_3: y_1 = x_2; y_2 = x_1; y_3 = x_3 \\
&\mathscr{R}_4: y_1 = x_2; y_2 = x_3; y_3 = x_1 \\
&\mathscr{R}_5: y_1 = x_3; y_2 = x_1; y_3 = x_2 \\
&\mathscr{R}_6: y_1 = x_3; y_2 = x_2; y_3 = x_1
\end{aligned}$$

因此在这个三维空间(x_1, x_2, x_3)中，存在着6个变换方程组，而且对于$y_1 < y_2 < y_3$有$6 = 3!$个不同的根。除了下面的情况，在y空间中是无解的

$$\begin{aligned}
\mathscr{R}_1: y_1 &= g_1(x_1,x_2,x_3) = x_1; x_1^{(1)} = \phi_1(y_1,y_2,y_3) = y_1 \\
y_2 &= h_1(x_1,x_2,x_3) = x_2; x_2^{(1)} = \varphi_1(y_1,y_2,y_3) = y_2 \\
y_3 &= q_1(x_1,x_2,x_3) = x_3; x_3^{(1)} = \theta_1(y_1,y_2,y_3) = y_3 \\
\mathscr{R}_4: y_1 &= g_4(x_1,x_2,x_3) = x_2; x_2^{(4)} = \phi_4(y_1,y_2,y_3) = y_1 \\
y_2 &= h_4(x_1,x_2,x_3) = x_3; x_3^{(4)} = \varphi_4(y_1,y_2,y_3) = y_2 \\
y_3 &= q_4(x_1,x_2,x_3) = x_1; x_1^{(4)} = \theta_4(y_1,y_2,y_3) = y_3 \\
\mathscr{R}_2: y_1 &= g_2(x_1,x_2,x_3) = x_1; x_1^{(2)} = \phi_2(y_1,y_2,y_3) = y_1 \\
y_2 &= h_2(x_1,x_2,x_3) = x_3; x_3^{(2)} = \varphi_2(y_1,y_2,y_3) = y_2 \\
y_3 &= q_2(x_1,x_2,x_3) = x_2; x_2^{(2)} = \theta_2(y_1,y_2,y_3) = y_3 \\
\mathscr{R}_5: y_1 &= g_5(x_1,x_2,x_3) = x_3; x_3^{(5)} = \phi_5(y_1,y_2,y_3) = y_1 \\
y_2 &= h_5(x_1,x_2,x_3) = x_1; x_1^{(5)} = \varphi_5(y_1,y_2,y_3) = y_2 \\
y_3 &= q_5(x_1,x_2,x_3) = x_2; x_2^{(5)} = \theta_5(y_1,y_2,y_3) = y_3 \\
\mathscr{R}_3: y_1 &= g_3(x_1,x_2,x_3) = x_2; x_2^{(3)} = \phi_3(y_1,y_2,y_3) = y_1 \\
y_2 &= h_3(x_1,x_2,x_3) = x_1; x_1^{(3)} = \varphi_3(y_1,y_2,y_3) = y_2 \\
y_3 &= q_3(x_1,x_2,x_3) = x_3; x_3^{(3)} = \theta_3(y_1,y_2,y_3) = y_3 \\
\mathscr{R}_6: y_1 &= g_6(x_1,x_2,x_3) = x_3; x_3^{(6)} = \phi_6(y_1,y_2,y_3) = y_1 \\
y_2 &= h_6(x_1,x_2,x_3) = x_2; x_2^{(6)} = \varphi_6(y_1,y_2,y_3) = y_2 \\
y_3 &= q_6(x_1,x_2,x_3) = x_1; x_1^{(6)} = \theta_6(y_1,y_2,y_3) = y_3
\end{aligned}$$

因为这些变换中的雅可比值都为1，对于3个有序的独立同分布的随机变量，式(5.2-11)可写为

$$\begin{aligned}
f_{Y_1Y_2Y_3}(y_1,y_2,y_3) &= \sum_{m=1}^{m=6} \frac{f_{X_1X_2X_3}(x_1^{(m)}, x_2^{(m)}, x_3^{(m)})}{|J_m|} \\
&= \sum_{m=1}^{m=6} f_X(x_1^{(m)}) f_X(x_2^{(m)}) f_X(x_3^{m)})
\end{aligned}$$

将求和表达式展开，并代入正确的解答可得最终结果为

$$\begin{aligned}
f_{Y_1Y_2Y_3}(y_1,y_2,y_3) &= f_X(y_1)f_X(y_2)f_X(y_3) + f_X(y_1)f_X(y_3)f_X(y_2) + f_X(y_2)f_X(y_1)f_X(y_3) \\
&\quad + f_X(y_2)f_X(y_3)f_X(y_1) + f_X(y_3)f_X(y_1)f_X(y_2) + f_X(y_3)f_X(y_2)f_X(y_1) \\
&= 3! f_X(y_1)f_X(y_2)f_X(y_3)
\end{aligned}$$

上面的结论在$y_1 < y_2 < y_3$时成立，否则$f_{Y_1Y_2Y_3}(y_1, y_2, y_3) = 0$。

此时，我们总结一般情况下的结论。如果已知n个独立同分布的连续随机变量，概率密度函数为$f_{X_1 \cdots X_n}(x_1, \cdots, x_n) = \prod_{i=1}^{n} f_X(x_i)$，且$-\infty < x_1, x_2, \cdots, x_n < \infty$。请考虑一个升序变换$Y_1 < Y_2 < \cdots < Y_n$，其中$i = 1, \cdots, n$，$Y_i \in \{X_1, X_2, \cdots, X_n\}$。此时可得

$$f_{Y_1\cdots Y_n}(y_1,\cdots,y_n)=\begin{cases}n!\prod_{i=1}^{n}f_X(y_i), & -\infty<y_1<y_2<\cdots<y_n<\infty\\ 0, & \text{其他}\end{cases} \tag{5.3-1}$$

如果$f_{Y_1\cdots Y_n}(y_1,\cdots,y_n)$是一个合法的概率密度函数，那么它的积分值一定是1，此时的积分是一个n重积分。

为了验证这个积分结果，我们先考虑$n=2$的情况。此时要将函数$2f_{Y_1Y_2}(y_1,y_2)=2f_{Y_1}(y_1)f_{Y_2}(y_2)$在区间$-\infty<y_1<y_2<\infty$上积分，需要将被积函数先在区间$-\infty<y_1<y_2$上积分，然后再在区间$-\infty<y_2<\infty$上积分，积分区间如图5.3-1所示。我们先在区间$-\infty<y_1<y_2$上，将变量$y_1$从$-\infty$积分到$y_2$；然后再在半空间上，将$y_2$从$-\infty$积分到$\infty$。

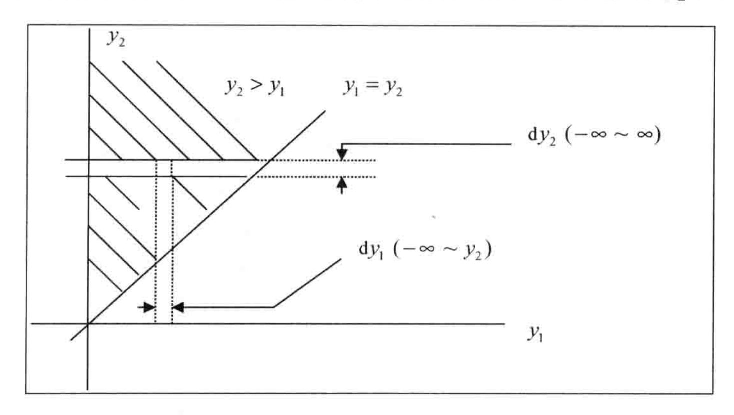


图5.3-1 两个有序随机变量的积分区域

可直接将上述结论扩展到n维情况。先将变量y_1从$-\infty$积分到y_2，然后再将变量y_2从$-\infty$积分到y_3，以此类推，最后将变量y_n从$-\infty$积分到∞。在这种情况下，我们在$-\infty<y_1<\cdots<y_n<\infty$所有的子空间上都进行了积分，最终的积分结果为

$$n!\int_{-\infty}^{\infty}F_X^{n-1}(y_n)f_X(y_n)\mathrm{d}y_n/(n-1)!=n!\int_{-\infty}^{\infty}F_X^{n-1}(y_n)\mathrm{d}F(y_n)/(n-1)!=F_X^n(y_n)|_{-\infty}^{\infty}=1$$

下面的推导将给出顺序统计学的一些基本结论。

面积随机变量的分布

首先给出面积随机变量的定义

$$Z_i\triangleq\int_{-\infty}^{Y_i}f_X(x)\mathrm{d}x,i=1,\cdots,n \tag{5.3-2}$$

其中$f_X(x)$是连续随机变量X的概率密度函数，$X_1,\cdots,X_n$是X的n个独立同分布的观测值。排序操作后得到有序随机变量Y_i，$\min(X_1,\cdots,X_n)\triangleq Y_1<Y_2<\cdots<Y_n\triangleq\max(X_1,\cdots,X_n)$，$i=1,\cdots,n$。我们用$Z_i$表示一个“面积随机变量”，因为随机变量$Z_i$表示函数$f_X(x)$到$Y_i$下方的面积，不过这种表示方法不太规范。显然，因为$Y_i$是随机变量，所以$Z_i$也是随机变量。实际上，可以把$Z_i$看成一个输入参数为随机变量的概率分布函数，因此也可以认为Z_i是一个随机的概率分布函数。因为$Y_1<\cdots<Y_n$，所以有$Z_1<\cdots<Z_n$，而且Z_i是关于Y_i的单调递增函数。我们考虑如下的变换：

$$z_i=\int_{-\infty}^{y_i}f_X(x)\mathrm{d}x=F_X(y_i),i=1,\cdots,n$$

其中 $F_X(x)$ 是 x 的一个连续的增函数，因此对于每个 x 都有唯一的逆像存在。这些方程的根为 $y_i^{(r)} = F_X^{-1}(z_i)$，$i = 1, \cdots, n$，(如图 5.3-2 所示)，雅可比值为

$$\begin{vmatrix} \frac{\partial z_1}{\partial y_1} & \cdots & \frac{\partial z_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_n}{\partial y_1} & \cdots & \frac{\partial z_n}{\partial y_n} \end{vmatrix} = \begin{vmatrix} f_X(y_1^{(r)}) & 0 & \cdots & 0 \\ 0 & f_X(y_2^{(r)}) & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & f_X(y_n^{(r)}) \end{vmatrix} = \prod_{i=1}^{n} f_X(y_i^{(r)}) \tag{5.3-3}$$

因此可得 Z_i 的概率密度函数为 $i = 1, \cdots, n$

$$\begin{aligned} f_{Z_1\cdots Z_n}(z_1, \cdots, z_n) &= n!\frac{\prod_{i=1}^{n} f_X(y_i^{(r)})}{\prod_{i=1}^{n} f_X(y_i^{(r)})} = n!,\ 0 < z_1 < z_2 < \cdots < z_n < 1 \\ &= 0,\ \text{其他} \end{aligned} \tag{5.3-4}$$

从这个结论可知，Z_i 的概率密度函数与 $f_X(x)$ 无关 $i = 1, \cdots, n$。当我们不知道某个变量的概率分布时，可以用式(5.3-4)的结论来估计各种参数，具体情况可参见例 5.3-5。

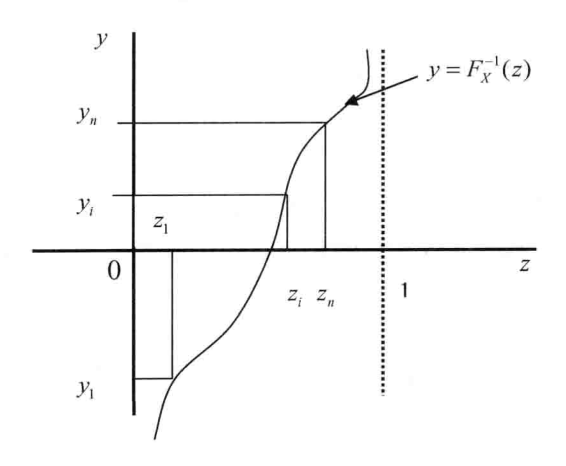


图 5.3-2　计算变换 $y = F_X^{-1}(z)$ 的根

例 5.3-1　[从最小值到最大值，$f_X(x)$ 下方的面积]我们希望计算 $f_X(x)$ 下方的面积，起点为 n 个观测样本中的最小值 $Y_1 = \min(X_1, \cdots, X_n)$，终点为最大值 $Y_n = \max(X_1, \cdots, X_n)$。我们用一个新的随机变量来表示这个面积

$$V_{1n} \triangleq \int_{Y_1}^{Y_n} f_X(x)\mathrm{d}x \tag{5.3-5}$$

注意到有下面的关系成立，因此需要根据 $f_{Z_1\cdots Z_n}(z_1, \cdots, z_n)$ 来计算 $f_{Z_1Z_n}(z_1, z_n)$。

$$V_{1n} = \int_{-\infty}^{Y_n} f_X(x)\mathrm{d}x - \int_{-\infty}^{Y_1} f_X(x)\mathrm{d}x = Z_n - Z_1 \tag{5.3-6}$$

将式(5.3-4)在 $z_2, z_3, \cdots, z_{n-1}$ 上进行积分，其中 $z_{i-1} < z_i < 1$，得到的结果为

$$\begin{aligned} f_{Z_1Z_n}(z_1, z_n) &= n(n-1)(z_n - z_1)^{n-2}, \quad 0 < z_1 < z_n < 1, n \geqslant 2 \\ &= 0,\ \text{其他} \end{aligned} \tag{5.3-7}$$

现在考虑两个新的随机变量 $V_{1n} \triangleq Z_n - Z_1$，$W \triangleq Z_n$。为了计算概率密度函数 $f_{VW}(v, w)$，需要

利用变换 $v = z_n - z_1$, $w = z_n$, $0 < v < w < 1$。这个变换的雅可比值为1，而且这个变换的唯一解为 $z_1^{(r)} = w - v$, $z_n^{(r)} = w$。因此可得

$$f_{V_{1n}W}(v,w) = n(n-1)v^{n-2}, \qquad 0 < w - v < w < 1, n \geqslant 2$$
$$= 0, \text{其他}$$

为了计算 V_{1n} 的概率密度函数，需要对 w 进行积分。根据两个不等式 $w - v > 0$ 和 $w < 1$ 可知，积分区域为一个三角形区域，如图5.3-3所示。

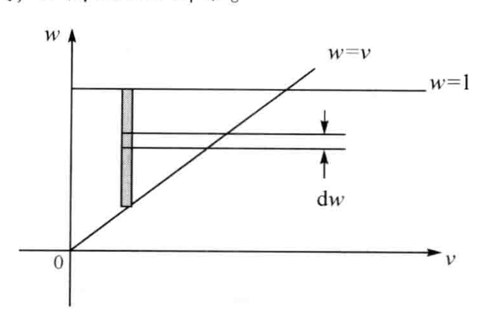


图5.3-3 计算 V_{1n} 概率密度函数用到的积分区域

根据 $f_{V_{1n}}(v) = n(n-1)v^{n-2}\int_v^1 \mathrm{d}w$ 可得

$$f_{V_{1n}}(v) = \begin{cases} n(n-1)v^{n-2}(1-v), & 0 < v < 1, n \geqslant 2 \\ 0, & \text{其他} \end{cases} \tag{5.3-8}$$

所得的概率密度函数是2.4节介绍的贝塔概率密度函数的一个特例，此时 $\alpha = n - 2$, $\beta = 1$。这个分布函数就是概率值，也就是最大值和最小值之间的面积，这个值是小于等于 v 的。最终结果如下：

$$F_{V_{1n}}(v) = \begin{cases} nv^{n-1} - (n-1)v^n, & 0 < v < 1 \\ 1, & v > 1 \\ 0, & v < 0 \end{cases} \tag{5.3-9}$$

图5.3-4所示为 n 取不同值时贝塔累计分布函数。

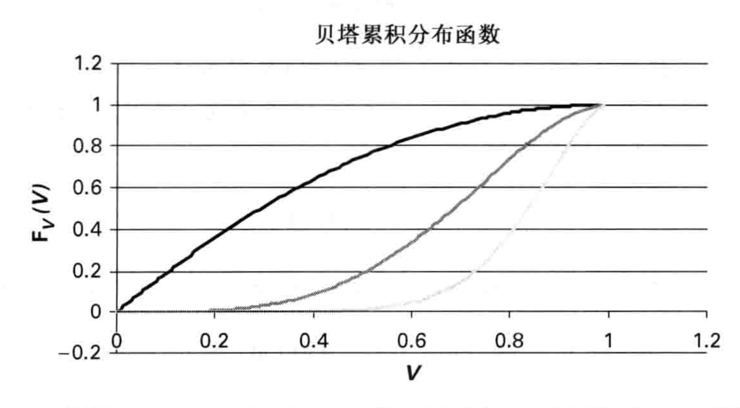


图5.3-4 $n=2$(最上方曲线)，$n=4$(中间曲线)$n=10$(最下方曲线)时的贝塔累积分布函数[参见式(5.3-9)]

例5.3-2 (任意两个有序随机变量之间的面积)我们拓展上个例题的结论，计算 $f_X(x)$ 在任意两个有序随机变量之间的面积，并不要求是在第一个和最后一个之间。我们用一个符

号来表示

$$V_{lm} \triangleq Z_m - Z_l = \int_{-\infty}^{Y_m} f_X(x)\mathrm{d}x - \int_{-\infty}^{Y_l} f_X(x)\mathrm{d}x = \int_{Y_l}^{Y_m} f_X(x)\mathrm{d}x, m > l \tag{5.3-10}$$

现考虑 $0 < Z_1 < Z_2 < Z_3 < 1$，其中 $f_{Z_1Z_2Z_3}(z_1, z_2, z_3) = 3!$，$0 < z_1 < z_2 < z_3 < 1$。首先计算随机变量 $V_{23} = Z_3 - Z_2$ 的概率密度函数 $f_{V_{23}}(v)$。因为这个函数只与 Z_2 和 Z_3 有关，所以必须从 $f_{Z_1Z_2Z_3}(z_1, z_2, z_3)$ 计算得到 $f_{Z_2Z_3}(z_2, z_3)$，计算如下所示：

$$f_{Z_2Z_3}(z_2, z_3) = \begin{cases} 3!\int_0^{z_2}\mathrm{d}z_1 = 3!z_2, & 0 < z_2 < z_3 < 1 \\ 0, & \text{其他} \end{cases}$$

为了从 $f_{Z_2Z_3}(z_2, z_3)$ 计算得到 $f_{V_{23}}(v)$，我们定义一个辅助随机变量 $B \triangleq Z_2$ 及其实现 β，还定义一个合适的方程组。在这种情况下，这个合适的方程组是 $v \triangleq z_3 - z_2$，$\beta \triangleq z_2$，方程组的根为 $z_2^{(r)} = \beta$，$z_3^{(r)} = v + \beta$。此时 $\beta \triangleq z_2$ 就成为一个辅助变量。此时，变换的雅可比值 $|J|$ 为 1，根据所谓的直接公式可得如下结果：

$$\begin{aligned} f_{BV_{23}}(\beta, v) &= f_{Z_2Z_3}(\beta, v+\beta)/|J| = 3!\beta, \quad 0 < \beta < 1 - v \\ &= 0, \text{其他} \end{aligned}$$

最后，将辅助变量 β 进行积分可得

$$\begin{aligned} f_{V_{23}}(v) &= 3!\int_0^{1-v}\beta\mathrm{d}\beta = \tfrac{3!(1-v)^2}{2!}, 0 < v < 1 \\ &= 0, \text{其他} \end{aligned} \tag{5.3-11}$$

可按照相同的思路，根据 $f_{Z_1Z_2Z_3}(z_1, z_2, z_3)$ 计算 $f_{V_{12}}(v)$。经计算可得，当 $0 < z_1 < z_2 < 1$ 时 $f_{Z_1Z_2}(z_1, z_2) = 3!(1 - z_2)$，其余区间为 $f_{Z_1Z_2}(z_1, z_2) = 0$。此时，根据变换 $v \triangleq z_2 - z_1$，$\beta \triangleq z_1$ 计算可得

$$\begin{aligned} f_{V_{12}}(v) &= 3!\int_0^{1-v}(1 - v - \beta)\mathrm{d}\beta = \tfrac{3!(1-v)^2}{2!}, 0 < v < 1 \\ &= 0, \text{其他} \end{aligned} \tag{5.3-12}$$

我们将详细推导过程留给读者。

在此给出一般情况的推导：设 V_{lm} 表示曲线 $f_X(x)$ 在 Y_l 和 Y_m 之间下方的面积(概率)，Y_1，Y_2，…，Y_n 表示概率密度函数 $f_X(x)$ 上有序的样本点。那么 V_{lm} 的概率密度函数可计算为

$$\begin{aligned} f_{V_{lm}}(v) &= \tfrac{n!}{(m-l-1)!(n-m+l)!}v^{m-l-1}(1-v)^{n-m+l}, 0 < v < 1 \\ &= 0, \text{其他} \end{aligned} \tag{5.3-13}$$

例 5.3-3 [$f_X(x)$ 在有序样本之间面积的数学期望]考虑面积随机变量 $0 < Z_1 < Z_2 < Z_3 < 1$，Z_i 的定义参见式(5.3-2)。我们希望计算 $E[Z_i]$，$i = 1, 2, 3$，验证 $E[Z_1] < E[Z_2] < E[Z_3]$ 是否成立。根据边缘概率密度函数 $f_{Z_i}(z_i)(i = 1,2,3)$ 的定义可计算

$$\begin{aligned} f_{Z_1}(z_1) &= \int_{z_1}^1\int_{z_1}^{z_3} f_{Z_1Z_2Z_3}(z_1, z_2, z_3)\mathrm{d}z_2\mathrm{d}z_3 = \tfrac{3!}{2!}(1 - z_1)^2 \\ f_{Z_2}(z_2) &= \int_{z_2}^1\int_0^{z_2} f_{Z_1Z_2Z_3}(z_1, z_2, z_3)\mathrm{d}z_1\mathrm{d}z_3 = 3!z_2(1 - z_2) \\ f_{Z_3}(z_3) &= \int_0^{z_3}\int_0^{z_2} f_{Z_1Z_2Z_3}(z_1, z_2, z_3)\mathrm{d}z_1\mathrm{d}z_2 = \tfrac{3!}{2!}z_3^2 \end{aligned}$$

根据上述结论可得

$$\begin{aligned} E[Z_1] &= \int_0^1 zf_{Z_1}(z)\mathrm{d}z = \tfrac{1}{4} = \tfrac{1}{3+1} \\ E[Z_2] &= \int_0^1 zf_{Z_2}(z)\mathrm{d}z = \tfrac{2}{4} = \tfrac{2}{3+1} \\ E[Z_3] &= \int_0^1 zf_{Z_3}(z)\mathrm{d}z = \tfrac{3}{4} = \tfrac{3}{3+1} \end{aligned}$$

根据这个结果可知：在一般情况下，$0 < Z_1 < Z_2 < \cdots < Z_n < 1$ 时，有下面的结论：

$$E[Z_i]=\frac{i}{n+1} \tag{5.3-14}$$

这个结论可以用数学归纳法来证明。

例 5.3-4 (有序样本之间面积的矩)考虑两个相邻有序样本之间的面积，即 $V_{i,\,i+1}=Z_{i+1}-Z_i$。$V_{i,\,i+1}$ 的概率密度函数参见式(5.3-13)，此时设 $m=i+1$，$l=i$，可得 $0<v<1$ 时 $f_{V_{i,\,i+1}}(v)=n(1-v)^{n-1}$，其余情况 $f_{V_{i,\,i+1}}(v)=0$，得到结果与 i 无关。根据这些结果可计算

$$E[V_{i,i+1}]=n\int_0^1 v(1-v)^{n-1}\mathrm{d}v=n\frac{\Gamma(2)\Gamma(n)}{\Gamma(n+2)}=\frac{1}{n+1}$$

其中伽马函数 $\Gamma(j)=(j-1)!$，$j=1,2,\cdots$，积分的计算查表可得(示例公式参见 p. 67 in *A Short Table of Integrals* by B. O. Peirce and R. M. Foster, Ginn and Company, New York, 1956)。这个积分也可以在网上很多地方查询到，包括 www.wolframalpha.com(在提示框里面输入“integral”)。

类似地可得

$$E[V_{i,i+1}^2]=n\int_0^1 v^2(1-v)^{n-1}\mathrm{d}v=n\frac{\Gamma(3)\Gamma(n)}{\Gamma(n+3)}=n\frac{2!(n-1)!}{(n+2)!}=\frac{2}{(n+2)(n+1)}$$

因此当 $n\gg 1$ 时，$\sigma_{V_{i,i+1}}^2=\dfrac{2}{(n+1)(n+2)}-\dfrac{1}{(n+1)^2}\approx\dfrac{1}{(n+1)^2}$。

为了计算方差 $\sigma_{Z_i}^2$，$i=1,\cdots,n$ 需要首先计算各个 $E[Z_i^2]$，$i=1,\cdots,n$，

$$\begin{aligned}
E[Z_1^2]&=n!\int_0^1\int_0^{z_n}\int_0^{z_{n-1}}\cdots\left(\int_0^{z_2}z_1^2\mathrm{d}z_1\right)\cdots\mathrm{d}z_{n-1}\mathrm{d}z_n=2\left((n+2)(n+1)\right)^{-1}\\
E[Z_2^2]&=n!\int_0^1\int_0^{z_n}\int_0^{z_{n-1}}\cdots\int_0^{z_3}z_2^2\int_0^{z_2}\mathrm{d}z_1\mathrm{d}z_2\cdots\mathrm{d}z_n=6\left((n+2)(n+1)\right)^{-1}\\
&\vdots\\
E[Z_n^2]&=n!\int_0^1 z_n^2\int_0^{z_n}\int_0^{z_{n-1}}\cdots\int_0^{z_3}\int_0^{z_2}\mathrm{d}z_1\mathrm{d}z_2\cdots\mathrm{d}z_n\\
&=n(n+1)/\left((n+2)(n+1)\right)^{-1}
\end{aligned}$$

可得一般结论为

$$E[Z_i^2]=\frac{i(i+1)}{(n+1)(n+2)},\quad i=1,\cdots,n$$

可根据式 $\sigma_i^2=E[Z_i^2]-E^2[Z_i]$ 计算方差如下：

$$\sigma_{Z_i}^2=\frac{i(i+1)}{(n+1)(n+2)}-\frac{i^2}{(n+1)^2}\approx\frac{i}{(n+1)^2},\quad n>>1$$

当 n 很大时，$\sigma_{Z_i}^2\approx E[Z_i]/(n+1)$。

例 5.3-5 (军靴尺码范围的估计)需要给新兵定做军用靴子，生产商需要知道这些军靴尺码的范围。假设测量得到了 n 个新兵的靴子尺寸的随机样本数据。请计算这个 n 的最小值，在这个范围内至少包含了 95% 的新兵靴子尺寸。

解 设 $\{X_i,\ i=1,\cdots,n\}$ 表示由 n 个新兵靴子尺寸组成的独立同分布的集合，这 n 个新兵是从总体中随机选取出来的，总体的概率密度函数 $f_X(x)$ 是未知的，设 $\{Y_i,\ i=1,\cdots,n\}$ 表示观测值的顺序统计量。根据 $V_{1n}=\int_{Y_1}^{Y_n}f_X(x)\mathrm{d}x$，我们需要计算 $P[V_{1n}\geqslant 0.95]=\delta$，在这里 δ 表示我们对 n 估计值可信度的一种度量。也就是说，在 $100\delta\,\%$ 的时间内，n 总是一个最小的样本数量，这个样本数量足够用来估计新兵靴子的尺寸。根据 $P[V_{1n}\leqslant 0.95]=1-\delta$ 与

式(5.3-9)可计算得到 $\delta = 0.95$ 时 $n = 93$，$\delta = 0.98$ 时 $n = 114$。计算结果用到了 Excel 的统计功能。注意：这里的结果与新兵数量是无关的。

5.4 期望向量和协方差矩阵[①]

定义 5.4-1 （列）向量 $\boldsymbol{X} = [X_1, \cdots, X_n]^{\mathrm{T}}$ 的数学期望是一个向量 $\boldsymbol{\mu}$（或者 $\bar{\boldsymbol{X}}$），该向量的各个元素 $\mu_1, \cdots, \mu_n$ 为

$$\mu_i \triangleq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_{\boldsymbol{X}}(x_1, \cdots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{5.4-1}$$

类似地

$$f_{X_i}(x_i) \triangleq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}x_1 \cdots \mathrm{d}x_{i-1}\,\mathrm{d}x_{i+1} \cdots \mathrm{d}x_n$$

X_i 的边缘概率密度函数可以写为

$$\mu_i = \int_{-\infty}^{\infty} x_i f_{X_i}(x_i)\,\mathrm{d}x_i, \qquad i = 1, \cdots, n$$

定义 5.4-2 一个实随机向量 $\boldsymbol{X}$ 的协方差矩阵 $\boldsymbol{K}$ 是向量 $[\boldsymbol{X} - \boldsymbol{\mu}][\boldsymbol{X} - \boldsymbol{\mu}]^{\mathrm{T}}$ 外积的数学期望，具体定义如下[②]：

$$\boldsymbol{K} \triangleq E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}}] \tag{5.4-2}$$

对于第 (i, j) 个分量

$$\begin{aligned} K_{ij} &\triangleq E[(X_i - \mu_i)(X_j - \mu_j)] \\ &= E[(X_j - \mu_j)(X_i - \mu_i)] \\ &= K_{ji}, \qquad i, j = 1, \cdots, n \end{aligned} \tag{5.4-3}$$

特别地，对于 $\sigma_i^2 \triangleq K_{ii}$，可以把 $\boldsymbol{K}$ 写成扩展形式

$$\boldsymbol{K} = \begin{bmatrix} \sigma_1^2 & & \cdots & & K_{1n} \\ & \ddots & & & \\ \vdots & & \sigma_i^2 & & \vdots \\ & & & \ddots & \\ K_{n1} & & \cdots & & \sigma_n^2 \end{bmatrix} \tag{5.4-4}$$

如果 $\boldsymbol{X}$ 是实向量，那么 $\boldsymbol{K}$ 的所有分量都是实数。因为 $K_{ij} = K_{ji}$，这种类型的实协方差矩阵也称为实对称的(r. s.)。这种类型的矩阵同时也可以归纳到更大的一类矩阵，称为埃尔米特矩阵(Hermitian)[③]。实对称矩阵具有很多有趣的性质，在接下来的章节我们会详细讨论。

对角元素 σ_i^2 其实是各个随机变量 X_i 的方差，$i = 1, \cdots, n$。协方差矩阵 $\boldsymbol{K}$ 和相关矩阵 $\boldsymbol{R}$ 是紧密联系的，相关矩阵定义如下：

---

① 本节内容需要读者稍微熟悉矩阵理论。

② 我们临时省略了均值，协方差和其他一些向量参数的下标，因为在这里很显然是在针对随机变量 $\boldsymbol{X}$。

③ 这种类型的矩阵是 $n \times n$ 维的，满足 $K_{ij} = K_{ji}^*$，该类型矩阵性质的详细讨论请参阅参考文献[5-2]。当 $\boldsymbol{X}$ 是复向量时，协方差一般都不是实对称的，但此时是埃尔米特共轭的。

$$\boldsymbol{R} \triangleq E[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}] \tag{5.4-5}$$

将式(5.4-2)展开可得

$$\boldsymbol{K} = \boldsymbol{R} - \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}$$

或者有

$$\boldsymbol{R} = \boldsymbol{K} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}} \tag{5.4-6}$$

对于实随机向量，其相关矩阵 $\boldsymbol{R}$ 也是实对称的，有的时候 $\boldsymbol{R}$ 也被称为自相关矩阵。随机向量经常按照是否相关，是否正交和是否独立来进行划分。

定义 5.4-3 对于两个 n 维随机向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$，均值向量分布为 $\boldsymbol{\mu}_X$ 和 $\boldsymbol{\mu}_Y$，如果二者外积的数学期望满足下面的等式：

$$E\{\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\} = \boldsymbol{\mu}_X\boldsymbol{\mu}_Y{}^{\mathrm{T}} \tag{5.4-7}$$

那么 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 是不相关的。如果

$$E\{\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\} = \boldsymbol{0} \qquad (\boldsymbol{0}\text{表示一个}n\times n\text{级的矩阵，所有元素为}0) \tag{5.4-8}$$

那么 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 是正交的。

对于正交的情况，$E\{X_iX_j\} = 0$，$0 \leqslant i, j \leqslant n$。此时二者内积的数学期望为0，即 $E[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}] = 0$。这也告诉我们，两个常规(非随机)向量正交的意思是 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0$。

最后，如果

$$f_{\boldsymbol{XY}}(\boldsymbol{x}, \boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x})f_{\boldsymbol{Y}}(\boldsymbol{y}) \tag{5.4-9}$$

那么 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 是独立的。

独立往往意味着不相关，但反之则不一定成立，5.6 节中介绍的多维高斯概率密度函数是一个特例。在实际情况中，很难证明两个随机向量是相互独立的。但是，在某个事先给定的置信度上，可利用统计实验确定二者相互关联的程度。

例 5.4-1 (几乎独立的随机变量)两个随机变量 X_1 和 X_2，当 $0 < x_1 \leqslant 1$，$0 < x_2 \leqslant 1$ 时联合概率密度函数为 $f_{X_1X_2}(x_1, x_2) = x_1 + x_2$，其余情况 $f_{X_1X_2}(x_1, x_2) = 0$。我们发现当 X_1 和 X_2 不是相互独立的，那么它们本质上是不相关的。为了证明这个结论，计算 $E[(X_1 - \mu_1)(X_2 - \mu_2)]$ 如下：

$$K_{12} = K_{21} = R_{21} - \mu_2\mu_1$$

首先可计算得到

$$\mu_1 = \mu_2 = \iint_S x(x+y)\,\mathrm{d}x\ \mathrm{d}y = 0.583$$

其中 $S = \{(x_1, x_2): 0 < x_1 \leqslant 1, 0 < x_2 \leqslant 1\}$。

接下来，再计算相关乘积

$$R_{12} = R_{21} \triangleq \iint_S xy(x+y)\,\mathrm{d}x\ \mathrm{d}y = 0.333$$

因此可得 $K_{12} = K_{21} = 0.333 - (0.583)^2 = -0.007$。同样地，我们还可以计算得到

$$\sigma_1^2 = \int_0^1 x^2(x+\tfrac{1}{2})\,\mathrm{d}x - (0.583)^2 = 0.4167 - 0.34 = 0.077$$

因此相关系数(归一化协方差)可计算为 $\rho = K_{12}/\sigma_1\sigma_2 = -0.091$。为了利用观测值 X_1 来预测

X_2，或者反过来用观测值 X_2 来预测 X_1，我们可能会设想这两个随机变量是不相关的。实际上，根据式(4.3-22)，例 4.3-4 的预测误差 ε 为 0.076，如果 X_1 和 X_2 确实是不相关的，那么预测误差应该是 0.077，此时的协方差矩阵 $\boldsymbol{K}$ 应该是

$$\boldsymbol{K}=\begin{bmatrix}0.077 & -0.007\\ -0.007 & 0.077\end{bmatrix}=0.077\begin{bmatrix}1 & -0.09\\ -0.09 & 1\end{bmatrix}$$

5.5　协方差矩阵的性质

因为协方差矩阵是实对称的，在此我们学习这种类型矩阵的一些性质。设 $\boldsymbol{M}$ 是一个 $n\times n$ 维的实对称矩阵。矩阵 $\boldsymbol{M}$ 的二次型为一个标量 $q(\boldsymbol{z})$，具体定义如下：

$$q(\boldsymbol{z})\triangleq\boldsymbol{z}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{z} \tag{5.5-1}$$

其中 $\boldsymbol{z}$ 是任意一个列向量。对于所有的 $\boldsymbol{z}$，如果下面的表达式成立：

$$\boldsymbol{z}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{z}\geqslant 0$$

那么 $\boldsymbol{M}$ 是一个半正定(p. s. d.)矩阵。如果表达式是严格成立的，即 $\boldsymbol{z}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{z}>0$，那么 $\boldsymbol{M}$ 是正定的(p. d.)。协方差矩阵 $\boldsymbol{K}$ 总是(至少是)半正定的，因为对于任意一个向量 $\boldsymbol{z}\triangleq[z_1, \cdots, z_n]^{\mathrm{T}}$，总有下面的结论成立：

$$\begin{aligned}0&\leqslant E\{[\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{\mu})]^2\}\\&=\boldsymbol{z}^{\mathrm{T}}E[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}]\boldsymbol{z}\\&=\boldsymbol{z}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{z}\end{aligned} \tag{5.5-2}$$

后面我们将证明：当矩阵 $\boldsymbol{K}$ 是满秩的，那么 $\boldsymbol{K}$ 是正定的。

现在，我们将给出线性代数[5-2, 第4章]中的一些定义和定理(大部分没有给出证明)。这些知识对于操作协方差矩阵非常有用。

定义 5.5-1　如果存在数 λ 和非零值 ϕ，使得特征方程 $\boldsymbol{M\phi}=\lambda\boldsymbol{\phi}$ 成立，那么 λ 就是 $n\times n$ 维矩阵 $\boldsymbol{M}$ 的特征值，列向量 $\boldsymbol{\phi}=[\phi_1, \phi_2, \cdots, \phi_n]^{\mathrm{T}}$ 称为特征向量。

特征向量总是需要被归一化，使得 $\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\phi}\triangleq\|\boldsymbol{\phi}\|^2=1$。

定理 5.5-1　当且仅当 $\det(\boldsymbol{M}-\lambda\boldsymbol{I})=0$ 时①，λ 为方阵 $\boldsymbol{M}$ 的特征值。

例 5.5-1　(特征值)对于下面给出的矩阵：

$$\boldsymbol{M}=\begin{bmatrix}4 & 2\\ 2 & 4\end{bmatrix}$$

根据定理 5.5-1 可计算 $\boldsymbol{M}$ 的特征值如下：

$$\det\begin{bmatrix}4-\lambda & 2\\ 2 & 4-\lambda\end{bmatrix}=(4-\lambda)^2-4=0$$

由此可得

$$\lambda_1=6, \qquad \lambda_2=2$$

① det 是 determinant(行列式)的简写，$\boldsymbol{I}$ 表示单位矩阵。

对应于 $\lambda_1=6$ 的(归一化)特征向量为

$$(\boldsymbol{M}-6\boldsymbol{I})\phi=0$$

将上式写成方程组的形式可得

$$\left.\begin{array}{l}-2\phi_1+2\phi_2=0\\-2\phi_1+2\phi_2=0\end{array}\right\}\Rightarrow\boldsymbol{\phi}_1=\frac{1}{\sqrt{2}}[1,1]^{\mathrm{T}}$$

双箭头符号“ $\Rightarrow$ ”意思是“意味着”。按照同样的步骤，可得 $\lambda_2=2$ 的特征向量如下：

$$\left.\begin{array}{l}2\phi_1+2\phi_2=0\\2\phi_1+2\phi_2=0\end{array}\right\}\Rightarrow\boldsymbol{\phi}_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$$

并不是所有 $n\times n$ 维的矩阵都有 n 个不同的特征值或 n 个特征向量。有的时候，一个矩阵的特征值个数少于 n，但该矩阵仍然拥有 n 个不同的特征向量。

定义 5.5-2 如果两个 $n\times n$ 维的矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是相似的，那么存在一个 $n\times n$ 的可逆矩阵 $\boldsymbol{T}$，$\det\boldsymbol{T}\neq 0$，使得下式成立：

$$\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}=\boldsymbol{B}\tag{5.5-3}$$

定理 5.5-2 如果一个 $n\times n$ 维的矩阵 $\boldsymbol{M}$ 与一个对角矩阵相似，当且仅当 $\boldsymbol{M}$ 存在 n 个线性独立的特征向量。

定理 5.5-3 设 $\boldsymbol{M}$ 是一个实对称矩阵，特征值为 λ_1，…，λ_n，那么 $\boldsymbol{M}$ 存在 n 个相互正交的单位特征向量 $\boldsymbol{\phi}_1$，…，$\boldsymbol{\phi}_n$ ①。

讨论 因为 $\boldsymbol{M}$ 具有 n 个相互正交(也是独立的)的单位特征向量，这类似于对角矩阵经过一个变换 $\boldsymbol{T}$ 后的对角矩阵 $\boldsymbol{\Lambda}$。请问 $\boldsymbol{T}$ 和 $\boldsymbol{\Lambda}$ 是什么？结果可由下面的重要定理计算得到。

定理 5.5-4 设 $\boldsymbol{M}$ 是一个实对称矩阵，特征值为 λ_1，…，λ_n，对角矩阵 $\boldsymbol{\Lambda}$ 的定义如下：

$$\boldsymbol{\Lambda}\triangleq\begin{bmatrix}\lambda_1 & & \boldsymbol{0}\\ & \ddots & \\ \boldsymbol{0} & & \lambda_n\end{bmatrix}$$

根据如下的变换可知，$\boldsymbol{M}$ 和 $\boldsymbol{\Lambda}$ 是相似的。

$$\boldsymbol{U}^{-1}\boldsymbol{M}\boldsymbol{U}=\boldsymbol{\Lambda}\tag{5.5-4}$$

其中矩阵 $\boldsymbol{U}$ 的各列由 $\boldsymbol{M}$ 相应②的单位正交特征向量 $\boldsymbol{\phi}_i$ 构成，$i=1$，…，n，即

$$\boldsymbol{U}=(\phi_1,\cdots,\phi_n)\tag{5.5-5}$$

进一步地，由于 $\boldsymbol{U}^{T}\boldsymbol{U}=\boldsymbol{I}$ (即 $\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{U}^{-1}$)，所以式(5.5-4)可以写为

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{U}=\boldsymbol{\Lambda}\tag{5.5-6}$$

讨论 矩阵 $\boldsymbol{U}$③ 如果满足 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}$，那么该矩阵就称为酉矩阵。酉矩阵具有保持范数的性质，即对于一个向量 $\boldsymbol{x}=[x_1,\cdots,x_n]^{\mathrm{T}}$，$\boldsymbol{x}$ 距离原点的欧氏距离为

$$||\boldsymbol{x}||\triangleq(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})^{1/2}$$

其中 $\|\boldsymbol{x}\|$ 称为 $\boldsymbol{x}$ 的模。对于变换 $\boldsymbol{y}=\boldsymbol{U}\boldsymbol{x}$，如果 $\boldsymbol{U}$ 是酉矩阵，那么

---

① 如果正交向量 $\boldsymbol{\phi}_i$ 满足 $\|\boldsymbol{\phi}_i\|=1$，那么该向量是标准正交的。

② 也就是说 ϕ_i 对应的是 λ_i，$i=1$，…，n。

③ 原文此处为 $\boldsymbol{M}$，疑有误——译者注。

$$||\boldsymbol{y}||^2=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{x}=||\boldsymbol{x}||^2$$

可知，经过变换 $\boldsymbol{y}=\boldsymbol{U}\boldsymbol{x}$ 后，新向量 $\boldsymbol{y}$ 与旧向量 $\boldsymbol{x}$ 与原点的距离都是相同的。

因为一个实随机向量的协方差矩阵 $\boldsymbol{K}$ 是实对称的，根据式(5.5-6)可知，只要 $\boldsymbol{U}$ 是已知的，那么 $\boldsymbol{K}$ 就可以被对角化。$\boldsymbol{U}$ 的各列正好是 $\boldsymbol{K}$ 的归一化特征向量，只要特征值已知就能得到特征向量。对协方差矩阵进行对角化是应用概率理论中一个非常重要的处理过程。这个过程可以将相关的随机变量转变为不相关的，而且如果是高斯的情况，还可以将相关的随机变量转换为独立的。

例 5.5-2　(随机向量的解相关)随机向量 $\boldsymbol{X}=[X_1,X_2,X_3]^{\mathrm{T}}$ 的协方差矩阵[①]如下所示：

$$\boldsymbol{K}_{\boldsymbol{XX}}=\begin{bmatrix}2&-1&1\\-1&2&0\\1&0&2\end{bmatrix}$$

请找出一个可逆的线性变换，$\boldsymbol{X}$ 通过该变换后产生新的随机向量 $\boldsymbol{Y}$，$\boldsymbol{Y}$ 的各分量是不相关的。

解　首先，通过求解方程 $\det(\boldsymbol{K}_{\boldsymbol{XX}}-\lambda\boldsymbol{I})=0$ 来计算特征值，可得 $\lambda_1=2$，$\lambda_2=2+\sqrt{2}$，$\lambda_3=2-\sqrt{2}$。接下来，再通过求解方程 $(\boldsymbol{K}_{\boldsymbol{XX}}-\lambda_i\boldsymbol{I})\boldsymbol{\phi}_i=\boldsymbol{0}$，$i=1,2,3$ 来计算三个特征向量，并将得到的特征向量进行归一化处理。将特征向量除以各自的模，即可得到归一化特征向量，具体为

$$\boldsymbol{\phi}_1=\left[0,\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right]^{\mathrm{T}}$$

$$\boldsymbol{\phi}_2=\left[\frac{1}{\sqrt{2}},-\frac{1}{2},\frac{1}{2}\right]^{\mathrm{T}}$$

$$\boldsymbol{\phi}_3=\left[\frac{1}{\sqrt{2}},\frac{1}{2},-\frac{1}{2}\right]^{\mathrm{T}}$$

现在，我们可以创建一个特征值矩量 $\boldsymbol{U}=[\boldsymbol{\phi}_1\quad\boldsymbol{\phi}_2\quad\boldsymbol{\phi}_3]$。通过转置操作得到的变换可使 $\boldsymbol{Y}$ 的各个分量都是不相关的

$$\boldsymbol{A}=\boldsymbol{U}^{\mathrm{T}}=\begin{bmatrix}0&\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{2}&\frac{1}{2}\\\frac{1}{\sqrt{2}}&\frac{1}{2}&-\frac{1}{2}\end{bmatrix}$$

通过变换 $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}$ 得到的各个分量为

$$Y_1=\frac{1}{\sqrt{2}}(X_2+X_3)$$

$$Y_2=\frac{1}{\sqrt{2}}X_1-\frac{1}{2}X_2+\frac{1}{2}X_3$$

$$Y_3=\frac{1}{\sqrt{2}}X_1+\frac{1}{2}X_2-\frac{1}{2}X_3$$

$\boldsymbol{Y}$ 的协方差矩阵为

---

① 我们给 $\boldsymbol{K}$ 加上下标，这是为了区分不同随机变量的协方差矩阵。

$$K_{YY}=\begin{bmatrix}2&0&0\\0&2+\sqrt{2}&0\\0&0&2-\sqrt{2}\end{bmatrix}$$

实际上，我们还可以对此题进一步分析。对 $\boldsymbol{Y}$ 的三个分量分别进行缩放，可以使得每个方差(即平均交流功率)都是相同的，这个处理过程也称为白化处理，接下来将会较为详细地讨论。如果将 Y_1 除以 $\dfrac{1}{\sqrt{\lambda_1}}$，$Y_2$ 除以 $\dfrac{1}{\sqrt{\lambda_2}}$，$Y_3$ 除以 $\dfrac{1}{\sqrt{\lambda_3}}$，此时得到的三个向量功率都是相同的。

如果 ϕ_1，…，ϕ_n 是实对称矩阵 $\boldsymbol{M}$ 的正交单位特征向量，那么对于如下的方程组：

$$\boldsymbol{M}\phi_1=\lambda_1\phi_1$$
$$\vdots$$
$$\boldsymbol{M}\phi_n=\lambda_n\phi_n$$

就可以写成如下紧凑的形式：

$$\boldsymbol{MU}=\boldsymbol{U}\Lambda \tag{5.5-7}$$

接下来的定理将会在实对称矩阵特征值和它的正定特性之间建立一个联系。

定理 5.5-5 一个实对称矩阵 $\boldsymbol{M}$ 是正定的，当且仅当它的所有特征值都为正。

证明 设 $\lambda_i>0$，$i=1$，…，n，对于任意一个向量 $\boldsymbol{x}$，通过线性变换 $\boldsymbol{x}\triangleq\boldsymbol{Uy}$ 后可得

$$\begin{aligned}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Mx}&=(\boldsymbol{Uy})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{Uy})\\&=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{MUy}\\&=\boldsymbol{y}^{\mathrm{T}}\Lambda\boldsymbol{y}\\&=\sum_{i=1}^{n}\lambda_i y_i^2>0\end{aligned} \tag{5.5-8}$$

当且仅当 $\boldsymbol{y}=0$ 时上式为 0。不过如果 $\boldsymbol{y}=0$，根据 $\boldsymbol{x}=\boldsymbol{Uy}$ 可得 $\boldsymbol{x}=0$。因此可以证明当所有的 $\lambda_i>0$ 时，矩阵 $\boldsymbol{M}$ 是正定的。反过来，我们还需要证明：如果 $\boldsymbol{M}$ 是正定的，那么所有的 λ_i 都是大于 0 的。因此，对于任意 $\boldsymbol{x}\neq 0$

$$0<\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Mx} \tag{5.5-9}$$

特别地，式(5.5-9)对于 ϕ_1，…，ϕ_n 都是成立的。

$$0<\phi_i^{\mathrm{T}}\boldsymbol{M}\phi_i=\lambda_i,\qquad i=1,\cdots,n$$

根据上式可知 $\lambda_i>0$，$i=1$，…，n。因此一个正定矩阵 $\boldsymbol{K}$ 的特征值都是正的。另外，行列式 det($\boldsymbol{K}$) 的值等于所有特征值的乘积，因此 det($\boldsymbol{K}$) >0。当 $\boldsymbol{K}$ 是满秩的，$\boldsymbol{K}$ 也是正定的。

白化变换

给定一个零均值的 $n\times 1$ 维的随机向量 $\boldsymbol{X}$，其协方差矩阵 $\boldsymbol{K}_{XX}$ 是正定的，我们希望找出一个变换 $\boldsymbol{Y}=\boldsymbol{CX}$ 使得 $\boldsymbol{K}_{YY}=\boldsymbol{I}$。矩阵 $\boldsymbol{C}$ 被称为白化变换，从 $\boldsymbol{X}$ 到 $\boldsymbol{Y}$ 的过程称为白化过程。设 $\boldsymbol{K}_{YY}$ 的 n 个特征向量和特征值分别用 ϕ_i 和 λ_i 来表示，$i=1$，…，n。此时可将特征方程 $\boldsymbol{K}_{XX}\phi_i=\lambda_i\phi_i$，$i=1,\cdots,n$，写成一个紧凑的形式 $\boldsymbol{K}_{XX}\boldsymbol{U}=\boldsymbol{U}\boldsymbol{\Lambda}$，其中 $\boldsymbol{U}\triangleq[\phi_1,\phi_2,\cdots,\phi_n]$，$\boldsymbol{\Lambda}\triangleq$ diag(λ_1，λ_2，…，λ_n)。因为 $\boldsymbol{K}_{YY}$ 是正定的，所以它的特征值都是正数，且存在矩阵 $\boldsymbol{\Lambda}^{-1/2}\triangleq$

$\text{diag}(1/\sqrt{\lambda_1}, 1/\sqrt{\lambda_2}, \cdots, 1/\sqrt{\lambda_n})$。对于变换 $\boldsymbol{Y} = \boldsymbol{CX} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}$，有

$\boldsymbol{K}_{\boldsymbol{YY}} = E[\boldsymbol{YY}^{\mathrm{T}}] = E[\boldsymbol{CXX}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}] = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^{\mathrm{T}}E[\boldsymbol{XX}^{\mathrm{T}}]\boldsymbol{U\Lambda}^{-1/2} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{XX}}\ \boldsymbol{U\Lambda}^{-1/2} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^{\mathrm{T}}(\boldsymbol{K}_{\boldsymbol{XX}}\boldsymbol{U})\boldsymbol{\Lambda}^{-1/2} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^{\mathrm{T}}(\boldsymbol{U\Lambda})\boldsymbol{\Lambda}^{-1/2} = \boldsymbol{\Lambda}^{-1/2}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})\boldsymbol{\Lambda\Lambda}^{-1/2} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda\Lambda}^{-1/2} = \boldsymbol{I}$，其中 $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{I}$

例5.5-3 （白化变换）在例5.5-2中，我们考虑一个随机向量 $\boldsymbol{X}$，其协方差矩阵和特征向量矩阵分别如下：

$$\boldsymbol{K}_{\boldsymbol{XX}} = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix} \boldsymbol{U} = \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/2 & 1/2 \\ 1/\sqrt{2} & 1/2 & -1/2 \end{bmatrix} = \boldsymbol{U}^{\mathrm{T}}$$

其中

$$\boldsymbol{\Lambda} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2+\sqrt{2} & 0 \\ 0 & 0 & 2-\sqrt{2} \end{bmatrix}$$

此时可得下面的变换即为白化变换：

$$\boldsymbol{Y} = \begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & (2+\sqrt{2})^{-1/2} & 0 \\ 0 & 0 & (2-\sqrt{2})^{-1/2} \end{bmatrix} \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/-2 & 1/2 \\ 1/\sqrt{2} & 1/2 & -1/2 \end{bmatrix} \boldsymbol{X}$$

白化变换在处理两个协方差矩阵对角化的时候非常有用①。

5.6 多维高斯(正态)分布

对于一般情况的 n 维高斯分布，看上去其数学表达式让人感到头痛，其实把它看成一维高斯概率密度函数的一种扩展就非常容易理解了。实际上，我们在4.3节就已经介绍过二维高斯概率密度函数，但在那里我们并没有将其扩展到一般情况。在这里考虑 n 维的情况，这样任意维的分布都可以看成 n 维的一种特例。如果已知 X 是一个(标量)高斯随机变量，其均值为 μ，方差为 σ^2，概率密度函数为

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

首先，考虑一个随机向量 $\boldsymbol{X} = [X_1, \cdots, X_n]^{\mathrm{T}}$，其各个分量 X_i 是相互独立的，且服从分布 $N(\mu_i, \sigma_i^2)$，$i = 1, \cdots, n$。那么 $\boldsymbol{X}$ 的概率密度函数就是各个独立分量 $X_1, \cdots, X_n$ 概率密度函数的乘积，即

$$\begin{aligned} f_{\boldsymbol{X}}(x_1, \cdots, x_n) &= \prod_{i=1}^{n} f_{X_i}(x_i) \\ &= \frac{1}{(2\pi)^{n/2}\sigma_1\cdots\sigma_n}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right] \end{aligned} \tag{5.6-1}$$

其中 μ_i 和 σ_i^2 分别是 X_i 的均值和方差，$i = 1, \cdots, n$。可将式(5.6-1)写成如下紧凑的形式：

① 这种对角化处理在模式识别中会用到，模式识别是应用概率论的一个分支。比如我们想对两组数据进行区分，如果数据是用对角化的协方差矩阵进行表征的，那么区分起来就比较容易。

$$f_{\boldsymbol{X}}(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}[\det(\boldsymbol{K}_{\boldsymbol{XX}})]^{1/2}}\exp[-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{XX}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})] \tag{5.6-2}$$

其中

$$\boldsymbol{K}_{\boldsymbol{XX}}\triangleq\begin{bmatrix}\sigma_1^2 & & 0\\ & \ddots & \\ 0 & & \sigma_n^2\end{bmatrix} \tag{5.6-3}$$

$\boldsymbol{\mu}=[\mu_1,\cdots,\mu_n]^{\mathrm{T}}$, $\det(\boldsymbol{K}_{XX})=\prod_{i=1}^{n}\sigma_i^2$。请注意 $\boldsymbol{K}_{XX}^{-1}$ 为如下表达式:

$$\boldsymbol{K}_{\boldsymbol{XX}}^{-1}=\begin{bmatrix}\sigma_1^{-2} & & 0\\ & \ddots & \\ 0 & & \sigma_n^{-2}\end{bmatrix}$$

另外，由于 X_i, $i=1,\cdots,n$ 是相互独立的，根据下面的推导可知，协方差矩阵 $\boldsymbol{K}_{XX}$ 是对角的。

$$E[(X_i-\mu_i)^2]\triangleq\sigma_i^2\quad i=1,\cdots,n \tag{5.6-4}$$

$$E[(X_i-\mu_i)(X_j-\mu_j)]=0,\quad i\neq j \tag{5.6-5}$$

接下来，请问：如果 $\boldsymbol{K}_{XX}$ 是一个正定矩阵(此时不要求是对角矩阵)会怎么样？将任意的正定协方差矩阵 $\boldsymbol{K}_{XX}$ 代入式(5.6-2)，此时仍然满足概率密度函数的要求吗？如果可以，把 $\boldsymbol{X}$ 称为正态随机向量，$f_{\boldsymbol{X}}(\boldsymbol{x})$ 为多维正态概率密度函数。为了证明 $f_{\boldsymbol{X}}(\boldsymbol{x})$ 确实是一个概率密度函数，我们必须证明下面两个性质：

$$f_{\boldsymbol{X}}(\boldsymbol{x})\geqslant 0 \tag{5.6-6a}$$

以及

$$\int_{-\infty}^{\infty}f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}=1 \tag{5.6-6b}$$

(用向量 $\mathrm{d}\boldsymbol{x}\triangleq\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$ 表示一个体积元)假设 $\boldsymbol{X}$ 始终是实的，即 $X_1,\cdots,X_n$ 是实的随机变量。将任意正定的协方差矩阵 $\boldsymbol{K}_{XX}$ 代入式(5.6-2)后满足式(5.6-6a)的要求，这个证明比较容易，我们将其留做练习。证明满足式(5.6-6b)比较复杂，接下来将给出证明。

式(5.6-6b)的证明 如果 $f_X(x)$ 如式(5.6-2)所示，那么 $\boldsymbol{K}_{XX}$ 是一个任意的正定协方差矩阵 利用 $\boldsymbol{z}\triangleq\boldsymbol{x}-\boldsymbol{\mu}$ 可以将式(5.6-2)改写为

$$f_{\boldsymbol{X}}(\boldsymbol{x})\triangleq\frac{1}{(2\pi)^{n/2}[\det(\boldsymbol{K}_{\boldsymbol{XX}})]^{1/2}}\phi(\boldsymbol{z})$$

其中

$$\phi(\boldsymbol{z})\triangleq\exp(-\tfrac{1}{2}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{XX}}^{-1}\boldsymbol{z}) \tag{5.6-7a}$$

同时 α 的定义如下：

$$\alpha\triangleq\int_{-\infty}^{\infty}\phi(\boldsymbol{z})\mathrm{d}\boldsymbol{z} \tag{5.6-7b}$$

此时可得

$$\int_{-\infty}^{\infty}f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}=\frac{\alpha}{(2\pi)^{n/2}[\det(\boldsymbol{K}_{\boldsymbol{XX}})]^{1/2}}$$

因此我们只需要通过估计 α 的值来证明式(5.6-6b)是成立的(或不成立)。

在对白化变换的学习中可知道，存在一个 $n\times n$ 的矩阵 $\boldsymbol{C}$，使得 $\boldsymbol{K}_{XX}=\boldsymbol{C}\,\boldsymbol{C}^{\mathrm{T}}$ 以及

$C^T K_{XX}^{-1} C = I$ 成立(I 是单位矩阵)。将如下的线性变换用于式(5.6-7a):

$$z = Cy \tag{5.6-8}$$

为了理解这个变换的意义,首先可以注意如下的等式:

$$z^T K_{XX}^{-1} z = y^T C^T K_{XX}^{-1} C y = ||y||^2 = \sum_{i=1}^{n} y_i^2$$

所以 $\phi(z)$ 的表达式为

$$\phi(z) = \prod_{i=1}^{n} \exp[-\tfrac{1}{2} y_i^2]$$

接下来,我们要用到高等微积分中的一个结论(参见 Kenneth Miller,[5 -5, p. 16]):对于式(5.6-8)给出的线性变换,其体积微元为如下表达式:

$$dz = |\det(C)| dy$$

其中 $dz \triangleq dz_1 \cdots dz_n$,$dy dy_1 \cdots dy_n$。因此式(5.6-7b)可变换为

$$\begin{aligned}
\alpha &= \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} y_i^2\right) dy_1 \cdots dy_n |\det(C)| \\
&= \left[\int_{-\infty}^{\infty} e^{-y^2/2}\, dy\right]^n |\det(C)| \\
&= [2\pi]^{n/2} |\det(C)|
\end{aligned}$$

但是,因为 $K_{XX} = CC^T$,$\det(K_{XX}) = \det(C)\det(C^T) = [\det(C)]^2$ 或者

$$|\det(C)| = |\det(K_{XX})|^{1/2} = (\det(K_{XX}))^{1/2}$$

因此,下面两个等式成立:

$$\alpha = (2\pi)^{n/2}[\det(K_{XX})]^{1/2}$$

$$\frac{\alpha}{[2\pi]^{n/2}[\det(K_{XX})]^{1/2}} = 1$$

式(5.6-6b)得证。

如下的等式其实满足了一个概率密度函数的所有要求,是一个多维正态分布的概率密度函数的表达式。

$$f_X(x) = \frac{1}{(2\pi)^{n/2}[\det(K_{XX})]^{1/2}} \exp[-\tfrac{1}{2}(x-\mu)^T K_{XX}^{-1}(x-\mu)] \tag{5.6-9}$$

现在请问:如果随机向量 Y 由如下表达式给出,请问 Y 的概率密度函数是什么?

$$Y \triangleq AX \tag{5.6-10}$$

在这里 A 是一个 $n \times n$ 的非奇异变换,答案由接下来的定理给出。

定理 5.6-1 设 X 是一个 n 维的正态随机向量,协方差矩阵 K_{XX} 是正定的,均值向量为 μ。设 A 为一个 n 维的非奇异线性变换,那么 $Y \triangleq AX$ 是一个 n 维的正态随机向量,协方差矩阵 $K_{YY} \triangleq A K_{XX} A^T$,均值向量为 $\beta \triangleq A\mu$。

证明 根据式(5.2-11)可知

$$f_Y(y) = \sum_{i=1}^{r} \frac{f_X(x_i)}{|J_i|} \tag{5.6-11}$$

其中 $\boldsymbol{Y}$ 是 $\boldsymbol{X}$ 的某个函数，即 $\boldsymbol{Y}=\boldsymbol{g}(\boldsymbol{X})\triangleq[\boldsymbol{g}_1(\boldsymbol{X}),\cdots,g_n(\boldsymbol{X})]^{\mathrm{T}}$，$\boldsymbol{x}_i$ 是方程组 $\boldsymbol{g}(\boldsymbol{x}_i)-\boldsymbol{y}=\boldsymbol{0}$ 的根，$i=1,\cdots,r$。J_i 是第 i 个根的雅可比值

$$J_i=\det\left(\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{x}}\right)\bigg|_{\boldsymbol{x}=\boldsymbol{x}_i}=\begin{vmatrix}\frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n}\\ \vdots & & \vdots\\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n}\end{vmatrix}_{\boldsymbol{x}=\boldsymbol{x}_i} \tag{5.6-12}$$

因为 $\boldsymbol{A}$ 是一个非奇异的线性变换，所以只有唯一解

$$\boldsymbol{Ax}-\boldsymbol{y}=\boldsymbol{0}\text{ 可得 }\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{y} \tag{5.6-13}$$

同样地

$$J_i=\det\left(\frac{\partial(\boldsymbol{Ax})}{\partial\boldsymbol{x}}\right)=\det(\boldsymbol{A}) \tag{5.6-14}$$

因此有

$$f_{\boldsymbol{Y}}(\boldsymbol{y})=\frac{1}{(2\pi)^{n/2}[\det(\boldsymbol{K}_{\boldsymbol{XX}})]^{1/2}|\det(\boldsymbol{A})|}\exp(-\tfrac{1}{2}[\boldsymbol{A}^{-1}\boldsymbol{y}-\boldsymbol{\mu}]^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{XX}}^{-1}[\boldsymbol{A}^{-1}\boldsymbol{y}-\boldsymbol{\mu}]) \tag{5.6-15}$$

可以把上面这个非常复杂的表达式写成式(5.6-9)那样的形式吗？首先，考虑如下等式：

$$[\det(\boldsymbol{K}_{\boldsymbol{XX}})]^{1/2}|\det(\boldsymbol{A})|=[\det(\boldsymbol{AK}_{\boldsymbol{XX}}\boldsymbol{A}^{\mathrm{T}})]^{1/2} \tag{5.6-16}$$

其次，将下式左边的第一项和最后一项左乘 $\boldsymbol{A}$，然后将 $\boldsymbol{A}$ 放入协方差矩阵的逆，凑齐等式可得

$$[\boldsymbol{A}^{-1}\boldsymbol{y}-\boldsymbol{\mu}]^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{XX}}^{-1}[\boldsymbol{A}^{-1}\boldsymbol{y}-\boldsymbol{\mu}]=[\boldsymbol{y}-\boldsymbol{A\mu}]^{\mathrm{T}}[\boldsymbol{AK}_{\boldsymbol{XX}}\boldsymbol{A}^{\mathrm{T}}]^{-1}[\boldsymbol{y}-\boldsymbol{A\mu}] \tag{5.6-17}$$

同时，因为 $\boldsymbol{A\mu}\triangleq\beta=E[\boldsymbol{Y}]$，$\boldsymbol{A}\boldsymbol{K}_{\boldsymbol{XX}}\boldsymbol{A}^{\mathrm{T}}=E[[\boldsymbol{Y}-\beta][\boldsymbol{Y}-\beta]^{\mathrm{T}}]=\boldsymbol{K}_{\boldsymbol{YY}}$，所以式(5.6-15)还可以写为

$$f_{\boldsymbol{Y}}(\boldsymbol{y})=\frac{1}{(2\pi)^{n/2}[\det(\boldsymbol{K}_{\boldsymbol{YY}})]^{1/2}}\exp[-\tfrac{1}{2}[\boldsymbol{y}-\beta]^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{YY}}^{-1}[\boldsymbol{y}-\beta]] \tag{5.6-18}$$

接下来就有一个非常自然的问题：可否将上述结论进行拓展？即经过更一般的(非平凡的)线性变换后，$\boldsymbol{Y}$ 还会是一个正态随机向量吗？接下来的定理将会给出解答，这也是定理5.6-1的一个推广。

定理5.6-2 设 $\boldsymbol{X}$ 是一个 n 维的正态随机向量，协方差矩阵 $\boldsymbol{K}_{\boldsymbol{XX}}$ 是正定的，均值向量为 $\boldsymbol{\mu}$。设 $\boldsymbol{A}_{mn}$ 是一个 $m\times n$ 的矩阵，秩为 m。那么对于如下的随机向量：

$$\boldsymbol{Y}=\boldsymbol{A}_{mn}\boldsymbol{X}$$

向量 $\boldsymbol{Y}$ 具有一个 m 维的正态概率密度函数，协方差矩阵 $\boldsymbol{K}_{\boldsymbol{YY}}$ 是正定的，均值向量为 $\boldsymbol{\beta}$，具体表达式如下：

$$\boldsymbol{K}_{\boldsymbol{YY}}\triangleq\boldsymbol{A}_{mn}\boldsymbol{K}_{\boldsymbol{XX}}\boldsymbol{A}_{mn}^{\mathrm{T}} \tag{5.6-19}$$

以及

$$\beta=\boldsymbol{A}_{mn}\boldsymbol{\mu} \tag{5.6-20}$$

该定理的证明与定理5.6-1的证明非常类似，可参阅参考文献[Miller 5-6, p.22]。

接下来，我们将给出一些有关正态随机变量变换的例题。

例 5.6-1 (独立性变换)设 $\boldsymbol{X}=[X_1, X_2]^{\mathrm{T}}$ 为零均值的正态随机向量，协方差矩阵 $\boldsymbol{K}_{XX}$ 为

$$\boldsymbol{K}_{XX}=\begin{bmatrix}3 & -1\\ -1 & 3\end{bmatrix}$$

请找出一个变换 $\boldsymbol{Y}=\boldsymbol{CX}$，使得 $\boldsymbol{Y}=[Y_1, Y_2]^{\mathrm{T}}$ 是一个正态随机向量，其各个分量的方差为1，且不相关(当然也就是独立的)。

解

$$E[\boldsymbol{YY}^{\mathrm{T}}]=E[\boldsymbol{CXX}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}]=\boldsymbol{CK}_{XX}\boldsymbol{C}^{\mathrm{T}}=\boldsymbol{I}$$

上式最右边成立的条件是：$\boldsymbol{Y}$ 的协方差矩阵 $\boldsymbol{K}_{YY}$ 满足下面的条件：

$$\boldsymbol{K}_{YY}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\triangleq\boldsymbol{I}$$

根据前面有关白化的讨论可知，对于矩阵 $\boldsymbol{C}$ 一定有 $\boldsymbol{C}=\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^{\mathrm{T}}$，其中 $\boldsymbol{\Lambda}^{-1/2}$ 是如下所示的正态化矩阵：

$$\boldsymbol{\Lambda}^{-1/2}\triangleq\begin{bmatrix}\lambda_1^{-1/2} & 0\\ 0 & \lambda_2^{-1/2}\end{bmatrix}$$

λ_i 是 $\boldsymbol{K}_{XX}$ 是特征值，$i=1, 2$，矩阵 $\boldsymbol{U}$ 的各列由 $\boldsymbol{K}_{XX}$ 的归一化特征向量构成($\boldsymbol{U}^{-1}=\boldsymbol{U}^{\mathrm{T}}$)。计算行列式 $\det(\boldsymbol{K}_{XX}-\lambda\boldsymbol{I})=0$ 可得 $\lambda_1=4$，$\lambda_2=2$。因此有

$$\boldsymbol{\Lambda}=\begin{bmatrix}4 & 0\\ 0 & 2\end{bmatrix},\qquad \boldsymbol{Z}=\boldsymbol{\Lambda}^{-1/2}=\begin{bmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{\sqrt{2}}\end{bmatrix}$$

接下来，求解下面的方程组：

$$(\boldsymbol{K}_{XX}-\lambda_1\boldsymbol{I})\boldsymbol{\phi}_1=0,\qquad ||\boldsymbol{\phi}_1||=1$$

$$(\boldsymbol{K}_{XX}-\lambda_2\boldsymbol{I})\boldsymbol{\phi}_2=0,\qquad ||\boldsymbol{\phi}_2||=1$$

可得 $\boldsymbol{\phi}_1=[1/\sqrt{2}, -1/\sqrt{2}]^{\mathrm{T}}$，$\boldsymbol{\phi}_2=[1/\sqrt{2}, 1/\sqrt{2}]^{\mathrm{T}}$，因此有

$$\boldsymbol{U}=(\boldsymbol{\phi}_1,\boldsymbol{\phi}_2)=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}$$

以及

$$\boldsymbol{C}=\frac{1}{\sqrt{2}}\begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$$

为了验证 $\boldsymbol{CK}_{XX}\boldsymbol{C}^{\mathrm{T}}$ 是否是一个单位协方差矩阵，可计算如下：

$$\frac{1}{2}\begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}3 & -1\\ -1 & 3\end{bmatrix}\begin{bmatrix}\frac{1}{2} & \frac{1}{\sqrt{2}}\\ -\frac{1}{2} & \frac{1}{\sqrt{2}}\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

有的时候，我们希望产生随机向量 $\boldsymbol{X}$ 一些相关的样本，此时协方差矩阵 $\boldsymbol{K}_{XX}$ 不是对角的。根据例5.6-1可知，对于如下的变换：

$$\boldsymbol{X}=\boldsymbol{C}^{-1}\boldsymbol{Y}\tag{5.6-21}$$

其中 $\boldsymbol{C} = \boldsymbol{Z}\boldsymbol{U}^{\mathrm{T}}$，该变换可产生一个正态随机向量，协方差矩阵为 $\boldsymbol{K}_{\boldsymbol{XX}}$。因此，有一种方法就是通过非相关的样本产生相关样本，对计算机产生的独立样本通过式(5.6-21)的变换即可。该变换就是例5.6-1的一个逆过程。

例5.6-2 （相关正态随机变量）联合正态随机 X_1 和 X_2 的联合概率密度函数表达式如下所示[参见式(4.3-27)，以及4.3节中的相关讨论]：

$$f_{X_1X_2}(x_1,x_2)=\frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}}\exp\left(\frac{-1}{2\sigma^2(1-\rho^2)}(x_1^2-2\rho x_1x_2+x_2^2)\right)$$

设相关系数 $\rho = -0.5$。请根据 X_1 和 X_2 找出两个联合正态随机变量 Y_1 和 Y_2，使得 Y_1 和 Y_2 是相互独立的。避免出现 $Y_1 = Y_2 = 0$ 的情况。

解 定义 $\boldsymbol{x} \triangleq (x_1, x_2)^{\mathrm{T}}$，$\boldsymbol{y} \triangleq (y_1, y_2)^{\mathrm{T}}$。根据 $\rho = -0.5$，可以把指数部分的二次项改写为

$$x_1^2+x_1x_2+x_2^2=\boldsymbol{x}^{\mathrm{T}}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\boldsymbol{x}=ax_1^2+(b+c)x_1x_2+\mathrm{d}x_2^2$$

其中 a，b，c，d 待确定。我们立即可知 $a = d = 1$，又因为矩阵是实对称的，可知 $b = c = 0.5$。我们可以把 $f_{X_1X_2}(x_1, x_2)$ 写为如下标准形式：

$$f_{X_1X_2}(x_1,x_2)=\frac{1}{2\pi[\det(\boldsymbol{K}_{\boldsymbol{XX}})]^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{x}^{\mathrm{T}}\ \boldsymbol{K}_{\boldsymbol{XX}}^{-1}\boldsymbol{x})\right)$$

其中

$$\boldsymbol{K}_{\boldsymbol{XX}}^{-1}=\frac{1}{\sigma^2(1-\rho^2)}\begin{bmatrix} a & b \\ c & d \end{bmatrix}=\frac{4}{3\sigma^2}\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

我们需要找出一个可以对角化 $\boldsymbol{K}_{\boldsymbol{XX}}^{-1}$ 的变换。这样就会使得 Y_1 和 Y_2 的联合概率密度函数拆开，即使得 Y_1 和 Y_2 相互独立。

因子 $4/3\sigma^2$ 影响了 $\boldsymbol{K}_{\boldsymbol{XX}}^{-1}$ 的特征值，但没有影响特征向量。为了计算 $\boldsymbol{K}_{\boldsymbol{XX}}^{-1}$ 的一系列正交的特征向量，我们只需要考虑如下的 $\tilde{\boldsymbol{K}}_{\boldsymbol{XX}}^{-1}$：

$$\tilde{\boldsymbol{K}}_{\boldsymbol{XX}}^{-1}\triangleq\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

此时可得 $\tilde{\lambda}_1 = 3/2$，$\tilde{\lambda}_2 = 1/2$。相应的归一化特征向量为 $\boldsymbol{\phi}_1 = (1/\sqrt{2})[1, 1]^{\mathrm{T}}$，$\boldsymbol{\phi}_2 = (1/\sqrt{2})[1, -1]^{\mathrm{T}}$，因此可得

$$\tilde{\boldsymbol{U}}\triangleq\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

计算对角协方差矩阵并不需要用 $1/\sqrt{2}$ 来归一化，因此我们在此省去这个因子，可得

$$\tilde{\boldsymbol{U}}^{\mathrm{T}}\tilde{\boldsymbol{K}}_{\boldsymbol{XX}}^{-1}\tilde{\boldsymbol{U}}=\mathrm{diag}(3,1)$$

因此所需的变换为①

$$\boldsymbol{Y}=\tilde{\boldsymbol{U}}^{\mathrm{T}}\boldsymbol{X}$$

也就是

---

① 在这里并不需要对协方差矩阵进行白化处理，这与例5.6-1是有区别的。同样地，对角化 $\boldsymbol{K}_{\boldsymbol{XX}}^{-1}$ 与对角化 $\boldsymbol{K}_{\boldsymbol{XX}}$ 实质上是等价的。

$$Y_1 = X_1 + X_2$$

$$Y_2 = X_1 - X_2$$

为了计算$f_{Y_1Y_2}(y_1, y_2)$，我们要用到第 3 章的式(3.4-21)

$$f_{Y_1Y_2}(y_1, y_2) = \sum_{i=1}^{n} f_{X_1X_2}(\boldsymbol{x}_i)/|J_i|$$

其中$\boldsymbol{x}_i \triangleq [x_1^{(i)}, x_2^{(i)}]^{\mathrm{T}}$表示方程组$\boldsymbol{y} - \tilde{\boldsymbol{U}}^{\mathrm{T}}\boldsymbol{x} = 0$的$n$个根，$J_i$表示雅可比值，$i = 1, \cdots, n$。对于$n = 1$，方程$\boldsymbol{y} - \tilde{\boldsymbol{U}}^{\mathrm{T}}\boldsymbol{x} = 0$只有唯一的解

$$x_1 = \frac{y_1 + y_2}{2}$$

$$x_2 = \frac{y_1 - y_2}{2}$$

雅可比值为

$$J = \det\left(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}\right) = \det\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = -2$$

因此可得

$$f_{Y_1Y_2}(y_1, y_2) = \frac{1}{2} f_{X_1X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{y_1^2}{2\sigma^2}\right] \cdot \frac{1}{\sqrt{2\pi\sigma'^2}} \exp\left[-\frac{y_2^2}{2\sigma'^2}\right]$$

其中$\sigma' \triangleq \sqrt{3}\sigma$。

例 5.6-1 和例 5.6-2 是下面定理的特例。

定理 5.6-3　设$\boldsymbol{X}$是零均值的正态随机向量，协方差矩阵$\boldsymbol{K}_{\boldsymbol{XX}}$是正定的。那么存在一个$n \times n$维非奇异矩阵$\boldsymbol{C}$，使得通过下列变换后：

$$\boldsymbol{Y} = \boldsymbol{C}^{-1}\boldsymbol{X}$$

$\boldsymbol{Y}$的各个分量$Y_1, \cdots, Y_n$是相互独立的，且方差为 1。

证明　设$\boldsymbol{C}^{-1} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^{\mathrm{T}}$，那么$\boldsymbol{K}_{\boldsymbol{XX}} = \boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}$，得证。

例 5.6-3　(一般瑞利分布)设$\boldsymbol{X} = [X_1, X_2, X_3]^{\mathrm{T}}$是一个正态随机向量，其协方差矩阵为

$$\boldsymbol{K}_{\boldsymbol{XX}} = \sigma^2\boldsymbol{I}$$

请计算$R_3 \triangleq \|\boldsymbol{X}\| = \sqrt{X_1^2 + X_2^2 + X_3^2}$的概率密度函数。

解　事件$\{R_3 \leqslant r\}$的概率就是R_3的概率分布函数$F_{R_3}(r)$，因此有

$$F_{R_3}(r) = \frac{1}{|(2\pi)^{3/2}[\sigma^2]^{3/2}|} \iiint_{\mathscr{F}} \exp\left[-\frac{1}{2\sigma^2}(x_1^2 + x_2^2 + x_3^2)\right] \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}x_3$$

其中$\mathscr{F} \triangleq \{(x_1, x_2, x_3): \sqrt{x_1^2 + x_2^2 + x_3^2} \leqslant r\}$。同时给出如下定义

$$x_1 \triangleq \zeta \cos\phi$$

$$x_2 \triangleq \zeta \sin\phi \cos\theta$$

$$x_3 = \zeta \sin\phi \sin\theta$$

上面的变换实现了从直角坐标系到球坐标系的转换。这个变换的雅可比值为 $\xi^2\sin\phi$。根据这个变换，可得 $F_{R_3}(r)$ 的表达式为 ($r > 0$)

$$F_{R_3}(r) = \frac{1}{(2\pi)^{3/2}(\sigma^2)^{3/2}} \int_0^r \int_{\theta=0}^{2\pi} \int_{\phi=0}^{\pi} \exp\left[-\frac{\zeta^2}{2\sigma^2}\right] \zeta^2 \sin\phi \, d\zeta \, d\theta \, d\phi$$

$$= \frac{4\pi}{(2\pi)^{3/2}[\sigma^2]^{3/2}} \int_0^r \zeta^2 \exp\left[-\frac{\zeta^2}{2\sigma^2}\right] d\zeta$$

对 $F_{R_3}(r)$ 关于 r 求导即可得到 $f_{R_3}(r)$ 的表达式如下：

$$f_{R_3}(r) = \frac{2r^2}{\Gamma(\frac{3}{2})[2\sigma^2]^{3/2}} \exp\left[-\frac{r^2}{2\sigma^2}\right] \cdot u(r) \tag{5.6-22}$$

其中 $u(r)$ 是单位阶跃函数，$\Gamma(3/2) = \sqrt{\pi}/2$。式(5.6-22)是对第 2 章中介绍的二维瑞利分布的一个拓展。对于更一般的 n 维瑞利分布，即 $R_n \triangleq \| \boldsymbol{X} \| = \sqrt{X_1^2 + \cdots + X_n^2}$，其概率密度函数表达式为

$$f_{R_n}(r) = \frac{2r^{n-1}}{\Gamma\left(\frac{n}{2}\right)[2\sigma^2]^{n/2}} \exp\left[-\frac{r^2}{2\sigma^2}\right] \cdot u(r) \tag{5.6-23}$$

证明式(5.6-23)需要用到 n 维的球坐标系。n 维球坐标系在许多数学参考书中都有介绍[5-5, p.9]。式(5.6-23)的理解很具挑战性，我们将它留给读者。

5.7 随机向量的特征函数

在式(4.7-1)中，定义了一个随机变量的特征函数

$$\Phi_X(\omega) \triangleq E[e^{j\omega X}]$$

可以将上面这个定义很自然地扩展到随机向量。设 $\boldsymbol{X} = [X_1, \cdots, X_n]^T$ 是一个 n 维的实随机向量，$\boldsymbol{\omega} = [\omega_1, \cdots, \omega_n]^T$ 是一个 n 维的实参数向量，那么可定义 $\boldsymbol{X}$ 的特征函数如下：

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega}) \triangleq E[e^{j\boldsymbol{\omega}^T \boldsymbol{X}}] \tag{5.7-1}$$

标量情况的定义也是显而易见的。对于连续随机向量，可按照下面等式来计算式(5.7-1)：

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(\boldsymbol{x}) e^{j\boldsymbol{\omega}^T \boldsymbol{x}} d\boldsymbol{x} \tag{5.7-2}$$

在式(5.7-2)中我们采用了 n 重积分，以及紧凑的表达式 $d\boldsymbol{x} = dx_1 \cdots dx_n$。如果 $\boldsymbol{X}$ 是离散的随机向量，那么可通过联合概率质量函数来计算 $\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})$

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega}) = \sum_{-\infty}^{+\infty} P_{\boldsymbol{X}}(\boldsymbol{x}) e^{j\boldsymbol{\omega}^T \boldsymbol{x}} \tag{5.7-3}$$

上面的这个求和表达式是 n 重求和。

对于上面两种情况，我们发现除了多了一个负号，$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})$ 就是 $f_{\boldsymbol{X}}(\boldsymbol{x})$ 或者 $P_{\boldsymbol{X}}(\boldsymbol{x})$ 的 n 维傅里叶变换。鉴于此，我们就可以通过 n 维傅里叶逆变换来得到概率密度函数(仍然要多出来一个负号)，也就是

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \Phi_{\boldsymbol{X}}(\boldsymbol{\omega}) e^{-j\boldsymbol{\omega}^T \boldsymbol{x}} d\boldsymbol{\omega} \tag{5.7-4}$$

用这个特征函数来计算联合矩非常有用，接下来将通过一些例题来进行讲解。

例5.7-1 (求混合矩)设$\boldsymbol{X}\triangleq[X_1, X_2, X_3]^{\mathrm{T}}$，$\boldsymbol{\omega}\triangleq[\omega_1, \omega_2, \omega_3]^{\mathrm{T}}$，请计算$E[\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{X}_3]$。

解 因为

$$\Phi_{\boldsymbol{X}}(\omega_1,\omega_2,\omega_3)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{\boldsymbol{x}}(x_1,x_2,x_3)\mathrm{e}^{\mathrm{j}[\omega_1x_1+\omega_2x_2+\omega_3x_3]}\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3$$

求偏导可得

$$\begin{aligned}&\frac{1}{j^3}\left.\frac{\partial^3\Phi_{\boldsymbol{X}}(\omega_1,\omega_2,\omega_3)}{\partial\omega_1\partial\omega_2\partial\omega_3}\right|_{\omega_1=\omega_2=\omega_3=0}\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2x_3f_{\boldsymbol{X}}(x_1,x_2,x_3)\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3\\&\triangleq E[X_1X_2X_3]\end{aligned}$$

对于任意阶的矩，只要它存在，都可以用例5.7-1的方法通过求偏导来计算得到，也就是

$$E[X_1^{k_1}\cdots X_n^{k_n}]=j^{-(k_1+\cdots+k_n)}\left.\frac{\partial^{k_1+\cdots+k_n}\Phi_{\boldsymbol{X}}(\omega_1,\cdots,\omega_n)}{\partial\omega_1^{k_1}\cdots\partial\omega_n^{k_n}}\right|_{\omega_1=\cdots=\omega_n=0}\tag{5.7-5}$$

写成下列的形式，并将各个乘积分量进行幂级数展开

$$E[\exp(\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{X})]=E\left[\exp\left(\mathrm{j}\sum_{i=1}^{n}\omega_iX_i\right)\right]=E\left[\prod_{i=1}^{n}\exp(\mathrm{j}\omega_iX_i)\right]$$

可以得到一个更复杂的公式

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})=\sum_{k_1=0}^{\infty}\cdots\sum_{k_n=0}^{\infty}E[X_1^{k_1}\cdots X_n^{k_n}]\frac{(\mathrm{j}\omega_1)^{k_1}}{k_1!}\cdots\frac{(\mathrm{j}\omega_n)^{k_n}}{k_n!}\tag{5.7-6}$$

不过这个公式可以很清晰地揭示X_i联合特征函数和联合矩之间的关系，$i=1, \cdots, n$。当然，式(5.7-6)仅在下面表达式对所有的非负整数$k_1, \cdots, k_n$都存在的时候才成立，即式(5.7-6)此时才收敛。

$$E[X_1^{k_1}\cdots X_n^{k_n}]$$

从式(5.7-2)，我们还可以得到特征函数的一些重要性质。

随机向量特征函数的一些性质

1. $|\Phi_{\boldsymbol{X}}(\omega)|\leqslant\Phi_{\boldsymbol{X}}(\mathbf{0})=1$。
2. $\Phi_{\boldsymbol{X}}^{*}(\omega)=\Phi_{\boldsymbol{X}}(-\omega)$（$*$表示共轭）。
3. 如果$\Phi_{\boldsymbol{X}}(\omega)$已知，那么$\boldsymbol{X}$各分量各子集的所有特征函数都可知。

我们用下面的这个例子来讲解最后一个性质。假设$\boldsymbol{X}=[X_1, X_2, X_3]^{\mathrm{T}}$的特征函数①为$\Phi_{\boldsymbol{X}}(\omega_1, \omega_2, \omega_3)=E[\exp\mathrm{j}(\omega_1X_1+\omega_2X_2+\omega_3X_3)]$，那么有

$$\Phi_{X_1X_2}(\omega_1,\omega_2)=\Phi_{X_1X_2X_3}(\omega_1,\omega_2,0)$$

$$\Phi_{X_1X_3}(\omega_1,\omega_3)=\Phi_{X_1X_2X_3}(\omega_1,0,\omega_3)$$

$$\Phi_{X_1}(\omega_1)=\Phi_{X_1X_2X_3}(\omega_1,0,0)$$

---

① 当$\boldsymbol{X}=[X_1, X_2, X_3]^{\mathrm{T}}$时，我们用$\Phi_{\boldsymbol{X}}(\cdot)$来代替$\Phi_{X_1X_2X_3}(\cdot)$。

在第4章曾经提到，在对独立随机变量求和的问题中特征函数非常有用。因此，假设 $\boldsymbol{X}=[X_1,\cdots,X_n]^{\mathrm{T}}$，$X_i$ 是独立的随机变量，边缘概率密度函数为 $f_{X_i}(x_i)$，$i=1,\cdots,n$。

$$Z=X_1+\cdots+X_n$$

上面这个和的概率密度函数可计算为

$$f_Z(z)=f_{X_1}(z)*\cdots*f_{X_n}(z) \tag{5.7-7}$$

但是，计算式(5.7-7)的这个 n 重卷积是非常麻烦的。用特征函数就可比较容易的计算 $f_Z(z)$，具体如下：

$$\begin{aligned}\Phi_{\boldsymbol{Z}}(\boldsymbol{\omega})&=E[\mathrm{e}^{\mathrm{j}\omega(X_1+\cdots+X_n)}]\\&=\prod_{i=1}^{n}E\left[\mathrm{e}^{\mathrm{j}\omega X_i}\right]\\&=\prod_{i=1}^{n}\Phi_{X_i}(\omega)\end{aligned} \tag{5.7-8}$$

上面推导中的第2行，是因为 $X_1,\cdots,X_n$ 是 n 个独立的随机变量，$i=1,\cdots,n$，因此 $Y_i=g_i(X_i)$ 也是 n 个独立的随机变量，所以有 $E[Y_1\cdots Y_n]=E[Y_1]\cdots E[Y_n]$。对式(5.7-8)进行傅里叶逆变换可得概率密度函数 $f_Z(z)$。如果 X_i 是离散的，上面的结论也是成立的，此时应该用到概率质量函数和离散傅里叶变换。接下来，我们将用一个例题来讲解如何利用这个方法来计算随机变量之和的概率密度函数。

例5.7-2 (独立同分布的泊松特征函数)设 $\boldsymbol{X}=[X_1,\cdots,X_n]^{\mathrm{T}}$，其中 X_i 是独立同分布的泊松分布随机变量，分布参数为 λ，$i=1,\cdots,n$。设 $Z=X_1+\cdots+X_n$。各个概率质量函数为

$$P_{X_i}(k)=\frac{\lambda^k\mathrm{e}^{-\lambda}}{k!} \tag{5.7-9}$$

特征函数为

$$\begin{aligned}\Phi_{X_i}(\omega)&=\sum_{k=0}^{\infty}\frac{\lambda^k\exp(\mathrm{j}\omega k)}{k!}\mathrm{e}^{-\lambda}\\&=\mathrm{e}^{\lambda(\exp(\mathrm{j}\omega)-1)}\end{aligned} \tag{5.7-10}$$

因此，根据独立性可得

$$\begin{aligned}\Phi_Z(\omega)&=\prod_{i=1}^{n}\mathrm{e}^{\lambda(\exp(\mathrm{j}\omega)-1)}\\&=\mathrm{e}^{n\lambda(\exp(\mathrm{j}\omega)-1)}\end{aligned} \tag{5.7-11}$$

比较式(5.7-11)和式(5.7-10)可知，$\Phi_Z(z)$ 就是概率质量函数 $P_Z(k)$ 的特征函数

$$P_Z(k)=\frac{\alpha^k\mathrm{e}^{-\alpha}}{k!},\qquad k=0,1\cdots \tag{5.7-12}$$

其中 $\alpha\triangleq n\lambda$。因此可知，$n$ 个独立同分布泊松随机变量之和也服从泊松分布，分布参数为 $n\lambda$。

高斯(正态)分布的特征函数

设 $\boldsymbol{X}$ 是一个实的高斯(正态)随机向量，协方差矩阵 $\boldsymbol{K}_{XX}$ 为非奇异的。根据定理5.6-3可知，$\boldsymbol{K}_{XX}$ 和 $\boldsymbol{K}_{XX}^{-1}$ 可写成如下形式：

$$\boldsymbol{K}_{XX}=\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}} \tag{5.7-13}$$

$$\boldsymbol{K}_{\boldsymbol{XX}}^{-1}=\boldsymbol{D}\boldsymbol{D}^{\mathrm{T}},\qquad \boldsymbol{D}\triangleq[\boldsymbol{C}^{\mathrm{T}}]^{-1} \tag{5.7-14}$$

其中 $\boldsymbol{C}$ 和 $\boldsymbol{D}$ 是非奇异的。上面的这个结论接下来会非常有用。根据定义，$\boldsymbol{X}$ 的特征函数为

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})=\frac{1}{(2\pi)^{n/2}[\det(\boldsymbol{K}_{\boldsymbol{XX}})]^{1/2}}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{XX}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)\cdot\exp(\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \tag{5.7-15}$$

根据下面的变换：

$$\boldsymbol{z}\triangleq\boldsymbol{D}^{\mathrm{T}}[\boldsymbol{x}-\boldsymbol{\mu}] \tag{5.7-16}$$

可知

$$\begin{aligned}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}&=[\boldsymbol{x}-\boldsymbol{\mu}]^{\mathrm{T}}\boldsymbol{D}\boldsymbol{D}^{\mathrm{T}}[\boldsymbol{x}-\boldsymbol{\mu}]\\&=[\boldsymbol{x}-\boldsymbol{\mu}]^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{XX}}^{-1}[\boldsymbol{x}-\boldsymbol{\mu}]\end{aligned} \tag{5.7-17}$$

这个变换的雅可比值为 $\det(\boldsymbol{D}^{\mathrm{T}})=\det(\boldsymbol{D})$。根据式(5.7-16)的变换可知，式(5.7-15)可变换为

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})=\frac{\exp(\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\mu})}{(2\pi)^{n/2}[\det(\boldsymbol{K}_{\boldsymbol{XX}})]^{1/2}|\det(\boldsymbol{D})|}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}\right)\cdot\exp(\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}(\boldsymbol{D}^{\mathrm{T}})^{-1}\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} \tag{5.7-18}$$

我们可通过下面的等式实现对被积函数的完全平方：

$$\begin{aligned}\exp[-\{\tfrac{1}{2}[\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}-2\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}(\boldsymbol{D}^{\mathrm{T}})^{-1}\boldsymbol{z}]\}]&=\exp(-\tfrac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}(\boldsymbol{D}^{\mathrm{T}})^{-1}(\boldsymbol{D})^{-1}\boldsymbol{\omega})\\&\quad\cdot\exp(-\tfrac{1}{2}||\boldsymbol{z}-\mathrm{j}\boldsymbol{D}^{-1}\boldsymbol{\omega}||^2)\end{aligned} \tag{5.7-19}$$

如果采用如下的结论，可以极大地简化式(5.7-18)和式(5.7-19)：(a)如果 $\boldsymbol{K}_{\boldsymbol{XX}}^{-1}=\boldsymbol{D}\boldsymbol{D}^{\mathrm{T}}$，那么 $\boldsymbol{K}_{\boldsymbol{XX}}=[\boldsymbol{D}^{\mathrm{T}}]^{-1}\boldsymbol{D}^{-1}$；(b) $\det(\boldsymbol{K}_{\boldsymbol{XX}}^{-1})=\det(\boldsymbol{D})\det(\boldsymbol{D}^{\mathrm{T}})=[\det(\boldsymbol{D})]^2=[\det(\boldsymbol{K}_{\boldsymbol{XX}})]^{-1}$，因此有 $|\det(\boldsymbol{D})|^{-1}=[\det(\boldsymbol{K}_{\boldsymbol{XX}})]^{1/2}$，以及如下的结论：

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})=\exp\left(\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{XX}}\boldsymbol{\omega}\right)\cdot\frac{1}{(2\pi)^{n/2}}\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{1}{2}||\boldsymbol{z}-\mathrm{j}\boldsymbol{D}^{-1}\boldsymbol{\omega}||^2}\mathrm{d}\boldsymbol{z}$$

等号右边的 n 重积分其实是 n 个一维高斯密度函数积分的乘积，每个密度函数的方差都是 1。因此积分结果为 $(2\pi)^{n/2}$，与分母 $(2\pi)^{-n/2}$ 抵消后可得正态随机向量的特征函数如下：

$$\Phi_{\boldsymbol{X}}(\boldsymbol{\omega})=\exp[\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\mu}-\tfrac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{XX}}\boldsymbol{\omega}] \tag{5.7-20}$$

其中 $\boldsymbol{\mu}$ 是均值向量，$\boldsymbol{\omega}=[\omega_1,\cdots,\omega_n]^{\mathrm{T}}$，$\boldsymbol{K}_{\boldsymbol{XX}}$ 是协方差矩阵。$\Phi_{\boldsymbol{X}}(\omega)$ 是多维复高斯变量 ω 的函数，因此高斯概率密度函数映射到了一个高斯特征函数。这个结果并不令人惊讶，因为我们知道一维傅里叶变换可以把一个高斯函数映射到另一个高斯函数。

类似地，一个随机向量 $\boldsymbol{X}=[X_1,\cdots,X_N]^{\mathrm{T}}$ 的联合矩量生成函数的定义为

$$\begin{aligned}M_{\boldsymbol{X}}(\boldsymbol{t})&\triangleq E\left[\exp\sum_{i=1}^{N}t_iX_i\right]\\&=\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_N=0}^{\infty}\frac{t_1^{k_1}}{k_1!}\cdots\frac{t_N^{k_N}}{k_N!}E\left[X_1^{k_1}X_2^{k_2}\cdots X_N^{k_N}\right]\end{aligned}$$

根据上面的等式，可仿照计算特征函数的方法来计算联合矩。

小结

在本章中，学习了多个随机变量的计算。我们发现，将多个随机变量整合成随机向量当成整体来进行处理更方便。我们还发现，对独立同分布的随机变量进行排序后可以推导出许多

统计结果，此时不用知道随机变量的具体分布。在5.3节中，推导出了面积随机变量(概率密度函数在有序样本之间的面积)的分布和它的各阶矩。在接下来的章节中，我们还看到有序随机变量在统计学的一个分支中扮演着重要的角色，称为非参数统计学或稳健统计学。在实际情况下，描述 n 个随机变量的联合概率分布往往是非常困难的，因此对于随机向量的情况，我们往往摒弃概率密度函数(或概率质量函数)，而改为采用那些非完全但更实用的性质。我们致力于那些低阶矩的特性，尤其是均值和协方差。因为在信号处理，通信理论，模式识别，多重回归分析以及工程和科学的其他领域，协方差矩阵都非常重要，我们还用到矩阵理论和线性代数的许多结论来揭示这些矩阵的性质。

我们还讨论了多维高斯(正态)分布以及随机向量的特征函数。通过线性变换，高斯随机向量映射成高斯随机向量。我们还介绍了如何推导一个变换，该变换可以把相关的随机变量变为不相关的。定义了随机向量的特征函数，该函数在计算各阶矩的过程中非常有用，而且在求解独立随机变量之和的问题中也非常有用，我们用一些例题对其进行了讲解。最后根据向量和矩阵方法，推导出了高斯随机向量的特征函数，该函数曲线形状类似于高斯分布。

习题

(带*号的习题带有一定的难度，需要花更多精力对教材进一步的学习。)

5.1 设 $f_{\boldsymbol{X}}(\boldsymbol{x})$ 由如下公式给出：

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = K\mathrm{e}^{-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Lambda}}u(\boldsymbol{x})$$

其中 $\boldsymbol{\Lambda} = (\lambda_1, \cdots, \lambda_n)^{\mathrm{T}}$，$\lambda_i > 0$ 对于所有 i 成立，$\boldsymbol{x} = (x_1, \cdots, x_n)^{\mathrm{T}}$，如果 $x_i \geqslant 0$，$u(\boldsymbol{x}) = 1$，$i = 1, \cdots, n$，其余情况 $u(\boldsymbol{x}) = 0$。K 是一个待确定的常数。请问 K 取多少才能确保 $f_{\boldsymbol{X}}(\boldsymbol{x})$ 是一个概率密度函数?

5.2 设 B_i 是 n 个不相交的穷举事件，$i = 1, \cdots, n$。试证明 $\boldsymbol{X}$ 的概率分布函数为如下形式：

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{i=1}^{n} F_{\boldsymbol{X}|B_i}(\boldsymbol{x}|B_i)P[B_i]$$

5.3 已知 $-\infty < x_i < \infty$，$i = 1, 2, \cdots, n$，设

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sigma_1\cdots\sigma_n}\exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{n}\left(\frac{x_i}{\sigma_i}\right)^2\right)\right\}$$

请证明所有的边缘概率密度函数都是高斯的。

5.4 设 X_1，X_2 和 X_3 为三个标准正态分布随机变量。设 $Y_i \in \{X_1, X_2, X_3\}$，$i = 1, 2, 3$。根据 X_i 的大小排序得到 $Y_1 < Y_2 < Y_3$。请计算联合概率密度函数 $f_{Y_1Y_2Y_3}(y_1, y_2, y_3)$。

5.5 计算习题5.4的概率分布函数 $F_Y(y)$，$i = 1, 2, 3$，并绘出函数曲线。

5.6 在5.4节中，我们介绍了随机变量 Z_1 和 Z_n。请证明 Z_1 和 Z_n 的联合概率密度函数为式(5.3-7)。

5.7 对于随机变量 $V_{1n} \triangleq Z_n - Z_1$，$W = Z_n$，请证明在区间 $0 < w - v < w < 1$，联合概率密度函数 $f_{V_{1n}W}(v, w) = n(n-1)v^{n-2}$，在其余区间 $f_{V_{1n}W}(v, w) = 0$。

5.8 根据上题的结果，试证明在区间 $0 < v < 1$，$f_{v_{1n}}(v) = n(n-1)v^{n-2}(1-v)$，$n \geqslant 2$，其余区间 $f_{v_{1n}}(v) = 0$。

5.9 请证明在函数 $f_{Z_1Z_2Z_3}(z_1, z_2, z_3) = 3!$ 在区间 $0 < z_1 < z_2 < z_3 < 1$ 下方的面积为1。

5.10 计算贝塔分布的累积分布函数，其中 $n = 2$，$\beta = 0$。

5.11 请推导式(5.3-11)，式(5.3-12)和式(5.3-13)。

5.12 利用Excel或其他类似的计算机程序来生成贝塔函数的概率分布函数曲线，n 取值分别为15，20，30，$\beta = 1$。当 $n \to \infty$ 时，请描述曲线的变化。

5.13 请推导式(5.3-14)。

5.14 请证明：平均而言，n 个有序随机变量将 $f_X(x)$ 下方的面积平均划分成了 $n+1$ 等份。

5.15 请证明：任何一个由两个向量外积生成的矩阵 $\boldsymbol{M}=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$，$\boldsymbol{M}$ 的秩至多为1。并请解释为什么 $\boldsymbol{R}\triangleq\boldsymbol{E}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$ 是满秩的？

5.16 设 $\{X_i,\ i=1,\cdots,n\}$ 是 X 的 n 个独立同分布的观测值，$\{Y_i,\ i=1,\cdots,n\}$ 是相应的顺序统计量，请证明 $F_{Y_n}(y)=F_X^n(y)$。

5.17 设 $\{X_i,\ i=1,\cdots,n\}$ 是 X 的 n 个独立同分布的观测值，$\{Y_i,\ i=1,\cdots,n\}$ 是相应的顺序统计量，请证明 $F_{Y_1}=1-(1-F_X(y))^n$。

5.18 设 $\{X_i,\ i=1,\cdots,n\}$ 是 X 的 n 个独立同分布的观测值，$\{Y_i,\ i=1,\cdots,n\}$ 是相应的顺序统计量。请证明 $F_{Y_r}(y)=\sum_{i=r}^{n}\binom{n}{i}F_X^i(y)\left[1-F_X(y)\right]^{n-1}$。

5.19 随机变量 X_1 和 X_2 的联合概率密度函数如下：

$$f_{X_1X_2}(x_1,x_2)=\begin{cases}\frac{1}{16}, & |x_1|<4,\quad 2<x_2<4\\ 0, & \text{其他}\end{cases}$$

请证明 X_1 和 X_2 独立且正交的。

5.20 设 $\boldsymbol{X}_i$ 是 n 个相互正交的随机向量，$i=1,\cdots,n$，请证明

$$E\left[\left|\left|\sum_{i=1}^{n}\boldsymbol{X}_i\right|\right|^2\right]=\sum_{i=1}^{n}\boldsymbol{E}[\|\boldsymbol{X}_i\|^2]$$

（提示：可以使用定义 $\|\boldsymbol{X}\|^2\triangleq\boldsymbol{X}^T\boldsymbol{X}$）。

5.21 设 $\boldsymbol{X}_i$ 是 n 个不相关的随机向量，$i=1,\cdots,n$，均值为 $\boldsymbol{\mu}_i\triangleq E[\boldsymbol{X}_i]$，请证明

$$\boldsymbol{E}\left[\left|\left|\sum_{i=1}^{n}(\boldsymbol{X}_i-\boldsymbol{\mu}_i)\right|\right|^2\right]=\sum_{i=1}^{n}E[\|\boldsymbol{X}_i-\boldsymbol{\mu}_i\|^2]$$

5.22 设 $\boldsymbol{X}_i$ 是 n 个不相关的随机向量，$E[\boldsymbol{X}_i]=\boldsymbol{\mu}_i$，$i=1,\cdots,n$。请证明

$$E\left[\sum_{i=1}^{n}(\boldsymbol{X}_i-\boldsymbol{\mu}_i)\sum_{j=1}^{n}(\boldsymbol{X}_j-\boldsymbol{\mu}_j)^{\mathrm{T}}\right]=\sum_{i=1}^{n}\boldsymbol{K}_i$$

其中 $\boldsymbol{K}_i\triangleq E[(\boldsymbol{X}_i-\boldsymbol{\mu}_i)(\boldsymbol{X}_i-\boldsymbol{\mu}_i)^{\mathrm{T}}]$。

5.23 请解释为什么下面的这些矩阵都不是随机实向量的协方差矩阵。

$$\begin{bmatrix}2 & -4 & 0\\ -4 & 3 & 1\\ 0 & 1 & 2\end{bmatrix}\quad\begin{bmatrix}4 & 0 & 0\\ 0 & 6 & 0\\ 0 & 0 & -2\end{bmatrix}\quad\begin{bmatrix}6 & 1+\mathrm{j} & 2\\ 1-\mathrm{j} & 5 & -1\\ 2 & -1 & 6\end{bmatrix}\quad\begin{bmatrix}4 & 6 & 2\\ 6 & 9 & 3\\ 9 & 12 & 16\end{bmatrix}$$

(a) (b) (c) (d)

5.24 (a) 设向量 $\boldsymbol{X}$ 的均值 $E[\boldsymbol{X}]=\boldsymbol{0}$，协方差矩阵 $\boldsymbol{K}_{\boldsymbol{XX}}$ 如下所示：

$$\boldsymbol{K}_{\boldsymbol{XX}}=\begin{bmatrix}3 & \sqrt{2}\\ \sqrt{2} & 4\end{bmatrix}$$

请找出一个线性变换 $\boldsymbol{C}$，使得 $\boldsymbol{Y}=\boldsymbol{C}\boldsymbol{X}$ 的协方差矩阵为

$$\boldsymbol{K}_{\boldsymbol{YY}}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

请问 $\boldsymbol{C}$ 是否是一个酉变换？

(b) 考虑如下的两个实对称矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{A}'$：

$$\boldsymbol{A}\triangleq\begin{bmatrix}a & b\\ b & c\end{bmatrix},\qquad \boldsymbol{A}'\triangleq\begin{bmatrix}a' & b'\\ b' & c'\end{bmatrix}$$

请证明：当 $a=c$，$a'=c'$ 时，乘积 $\boldsymbol{A}\boldsymbol{A}'$ 是实对称的。更一般地，如果 $\boldsymbol{A}$ 和 $\boldsymbol{A}'$ 是任意的实对称矩阵且满足 $\boldsymbol{A}\boldsymbol{A}'=\boldsymbol{A}'\boldsymbol{A}$，那么 $\boldsymbol{A}\boldsymbol{A}'$ 是实对称的。

提示：(K. Fukunaga [5-8, p. 33].) 设 $\boldsymbol{K}_1$ 和 $\boldsymbol{K}_2$ 是正定的协方差矩阵，利用下面的等式：

$$\boldsymbol{K}=a_1\boldsymbol{K}_1+a_2\boldsymbol{K}_2,\qquad \text{其中}\, a_1,a_2>0$$

5.25 变换 $\boldsymbol{A}$ 具有下列性质：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{A}=\boldsymbol{I}\qquad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}_1\boldsymbol{A}=\boldsymbol{\Lambda}^{(1)}=\operatorname{diag}(\lambda_1^{(1)},\cdots,\lambda_n^{(1)})$$

(a)证明 $\boldsymbol{A}$ 满足下面的等式：

$$\boldsymbol{K}^{-1}\boldsymbol{K}_1\boldsymbol{A}=\boldsymbol{A}\boldsymbol{\Lambda}^{(1)}$$

(b)证明 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}_2\boldsymbol{A}\triangleq\boldsymbol{\Lambda}^{(2)}$ 也是对角矩阵，即 $\boldsymbol{\Lambda}^{(2)}\triangleq\operatorname{diag}(\lambda_1^{(2)},\cdots,\lambda_n^{(2)})$。

(c)证明 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}_1\boldsymbol{A}$ 和 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{K}_2\boldsymbol{A}$ 具有相同的特征向量。

(d)证明 $\boldsymbol{\Lambda}^{(2)}$ 和 $\boldsymbol{\Lambda}^{(1)}$ 的特征值具有下面的关系：

$$\lambda_i^{(2)}=\frac{1}{a_2}[1-a_1\lambda_i^{(1)}]$$

同时证明 $\boldsymbol{\Lambda}^{(2)}$ 特征值是 $\boldsymbol{\Lambda}^{(1)}$ 特征值的逆序。

5.26 (J. A. McLaughlin [5-9].)设 m 个向量 $\boldsymbol{X}_i=(X_{i1},\cdots,X_{in})^{\mathrm{T}}$, $i=1,\cdots,m$, $n>m$，以及一个 $n\times n$ 的矩阵 $\boldsymbol{S}=\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}}$。

(a)请证明：$\boldsymbol{S}$ 可以写成如下的形式：

$$\boldsymbol{S}=\frac{1}{m}\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}$$

其中 $\boldsymbol{W}\triangleq(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_m)$。

(b)矩阵 $\boldsymbol{S}$ 的最大秩是多少?

(c)设 $\boldsymbol{S}'\triangleq\frac{1}{m}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}$，请问 $\boldsymbol{S}'$ 是多少维的? 请证明：S 的前 m 个非零特征值可通过下面的等式计算得到：

$$\boldsymbol{S}'\boldsymbol{\Phi}=\boldsymbol{\Phi}\boldsymbol{\Lambda}$$

其中 $\boldsymbol{\Phi}$ 是 $\boldsymbol{S}'$ 的特征向量矩阵，$\boldsymbol{\Lambda}$ 是特征值矩阵。同时请问 $\boldsymbol{S}$ 与 $\boldsymbol{S}'$ 特征向量特征值之间有什么联系?

(d)请问通过 $\boldsymbol{S}'$ 来计算特征向量比通过 $\boldsymbol{S}$ 有何优势?

5.27 (a)设 $\boldsymbol{K}$ 是一个 $n\times n$ 的协方差矩阵，$\Delta\boldsymbol{K}$ 是一个实对称的扰动矩阵。设 λ_i 是 $\boldsymbol{K}$ 的特征值，ϕ_i 是相应的特征向量，$i=1,\cdots,n$。请证明 $\boldsymbol{K}+\Delta\boldsymbol{K}$ 特征值的一阶近似 λ_i' 为

$$\lambda_i'=\phi_i^{\mathrm{T}}(\boldsymbol{K}+\Delta\boldsymbol{K})\phi_i,\qquad i=1,\cdots,n$$

(b)请证明特征向量的一阶近似为，

$$\Delta\phi_i=\sum_{j=1}^{n}b_{ij}\phi_j$$

其中 $b_{ij}=\phi_j^{\mathrm{T}}\Delta\boldsymbol{K}\phi_i/(\lambda_i-\lambda_j)$, $i\neq j$, $b_{ii}=0$。

5.28 设 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$ 是实对称矩阵 $\boldsymbol{M}$ 的特征值。当 $i\geqslant 2$ 时，设 $\phi_1,\phi_2,\cdots,\phi_{i-1}$ 是相互正交的单位特征向量，分别对应于 $\lambda_1,\cdots,\lambda_{i-1}$。请证明：当 $\|\boldsymbol{u}\|=1$, $\boldsymbol{u}^{\mathrm{T}}\phi_1=\cdots=\boldsymbol{u}^{\mathrm{T}}\phi_i-1=0$ 时 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{u}$ 的最大值为 λ_i，即 $\lambda_i=\max(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{u})$。

5.29 设随机向量 $\boldsymbol{X}=(X_1,X_2,X_3)^{\mathrm{T}}$ 的均值 $\boldsymbol{\mu}\triangleq E[\boldsymbol{X}]$ 为

$$\boldsymbol{\mu}=(5,-5,6)^{\mathrm{T}}$$

方差阵为

$$\boldsymbol{K}=\begin{bmatrix}5&2&-1\\2&5&0\\-1&0&4\end{bmatrix}$$

请计算如下变换的均值和方差阵：

$$Y=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}+B$$

其中 $\boldsymbol{A}=(2,-1,2)^{\mathrm{T}}$, $B=5$。

5.30 联合正态随机向量 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$ 的联合概率密度函数表达式为

$$f_{X_1X_2}(x_1,x_2)=\frac{2}{\pi\sqrt{7}}\exp[-\tfrac{8}{7}(x_1^2+\tfrac{3}{2}x_1x_2+x_2^2)]$$

请找出一个非平凡变换 $\boldsymbol{A}$，使得 Y_1 和 Y_2 相互独立，并计算 Y_1 和 Y_2 的联合概率密度函数。

$$\begin{pmatrix}Y_1\\Y_2\end{pmatrix}=\boldsymbol{A}\begin{pmatrix}X_1\\X_2\end{pmatrix}$$

5.31 如果 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}$ 的均值为 $\boldsymbol{\mu}=(\mu_1,\cdots,\mu_n)^{\mathrm{T}}$，方差为

$$\boldsymbol{K}=\{K_{ij}\}_{n\times n}$$

对于变量随机变量 Y

$$Y\triangleq p_1X_1+\cdots+p_nX_n$$

请证明 Y 的均值和方差分别为如下表达式：

$$E[Y]=\sum_{i=1}^{n}p_i\mu_i$$

$$\sigma_Y^2=\sum_{i=1}^{n}\sum_{j=1}^{n}p_ip_jK_{ij}$$

5.32 请计算 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}$ 的联合特征函数，其中 X_i 是独立同分布的柯西随机变量

$$f_{X_i}(x)=\frac{\alpha}{\pi(x^2+\alpha^2)}$$

根据得到的结果计算 $Y=\sum_{i=1}^{n}X_i$ 的概率密度函数。

5.33 请计算 $\boldsymbol{X}=(X_1,\cdots,X_n)^{\mathrm{T}}$ 的联合特征函数，其中 X_i 是独立同分布的二项分布随机变量 $i=1,\cdots,n$。请根据得到的结果计算 $Y=\sum_{i=1}^{n}X_i$ 的概率质量函数。

5.34 设 $\boldsymbol{X}=(X_1,\cdots,X_4)^{\mathrm{T}}$ 是一个高斯随机向量，均值 $E[\boldsymbol{X}]=0$。请证明下式成立：

$$E[X_1X_2X_3X_4]=K_{12}K_{34}+K_{13}K_{24}+K_{14}K_{23}$$

其中 K_{ij} 是 $\boldsymbol{X}$ 的协方差矩阵 $\boldsymbol{K}=\{K_{ij}\}_{4\times4}$ 的元素。

5.35 设 X_1, X_2, X_3 的联合概率密度函数为 $f_{\boldsymbol{X}}(x_1,x_2,x_3)=2/3\cdot(x_1+x_2+x_3)$，定义域为 $S=\{(x_1,x_2,x_3):0<x_i\leqslant1, i=1,2,3\}$，在其余区间为 $f_{\boldsymbol{X}}(x_1,x_2,x_3)=0$。请计算协方差矩阵，并请证明 X_1, X_2, X_3 虽然不是相互独立的，但却不相关。

5.36 设 X_1, X_2 是联合高斯的，均值为零，协方差矩阵如下：

$$\boldsymbol{K}=\begin{bmatrix}2&-1.5\\-1.5&2\end{bmatrix}$$

请找出 $\boldsymbol{X}=(X_1X_2)^{\mathrm{T}}$ 的一个白化变换。编写一个 MATLAB 程序来绘出 x_1 和 x_2 的散点图，x_1 和 x_2 分别是 X_1 和 X_2 的实现。同时也画出白化处理后的散点图。仿真的点数分别选择 100 和 1000。

5.37 （线性变换）设 $Y_k=\sum_{j=1}^{n}a_{kj}X_j$，$k=1,\cdots,n$，其中 a_{kj} 是实常数，矩阵 $\boldsymbol{A}=[a_{ij}]_{N\times N}$ 是非奇异的，$\{X_j\}$ 是随机变量。设 $\boldsymbol{B}=\boldsymbol{A}^{-1}$。请证明 $\boldsymbol{Y}$ 的概率密度函数 $f_{\boldsymbol{Y}}(y_1,\cdots,y_n)$ 为如下表达式：

$$f_{\boldsymbol{Y}}(y_1,\cdots,y_n)=|\det\boldsymbol{B}|f_{\boldsymbol{X}}(x_1^*,\cdots,x_n^*),\quad x_i^*=\sum_{k=1}^{n}b_{ik}y_k,\quad i=1,\cdots,n$$

5.38 （辅助变量）设 $Y_1=\sum_{i=1}^{n}X_i$，$Y_2=\sum_{i=2}^{n}X_i$。请计算联合概率密度函数 $f_{Y_1Y_2}(y_1,y_2)$。提示：引入辅助变量 $Y_k=\sum_{i=k}^{n}X_i$，$k=3,\cdots,n$，并对每个辅助随机变量进行积分。证明 $f_{\boldsymbol{Y}}(y_1,\cdots,y_n)=f_{\boldsymbol{X}}(y_1-y_2,\cdots,y_{n-1}-y_n,y_n)$（本题以及上题来自 *Probability and Stochastic Processes for Engineers*，C. W. Helstrom，Macmillan，1984，改编自该书例 4.9，p.190）。

参考文献

5-1 W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edition. New York: John Wiley, 1971.

5-2 W. B. Davenport, Jr., *Probability and Random Processes*, New York: McGraw-Hill, p. 99.

5-3 A. Papoulis, and S. U. Pillai *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, 4th Ed, 2002.

5-4 B. Saleh, *Photoelectron Statistics.* New York: Springer-Verlag, 1978, Chapter 5.

5-5 R. G. Gallagher, *Information Theory and Reliable* Communications. New York: John Wiley, 1968.

5-6 P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Part 1. New York: McGraw-Hill, 1953, p. 279.

5-7 G. A. Korn and T. S. Korn, *Mathematical Handbook for Scientists and Engineers.* New York: McGraw-Hill, 1961.

5-8 W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd edition. New York: John Wiley, 1957.

5-9 K. Fukunaga, Introduction to Statisticak Pattern recognition. 2nd edition. New York: Academic 1960.

5-10 J. A. McLaughlin and J. Raviv, "Nth Order Autocorrelations in Pattern Recognition, Information and Control", 12, pp. 121-142, Chapter 2, 1968 .

补充文献

Cooper, G. R. and C. D. McGillem, *Probabilistic Methods of Signal and System Analysis*, 3rd edition. New York: Holt, Rinehart and Winston, 1999.

Peebles, P. Z. Jr., *Probability, Random Variables, and Random Signal Principles*, 4th edition. New York: McGraw-Hill, 2001.

Leon-Garcia, A., *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd edition. Reading, MA: Prentice Hall, 2008.

Helstrom, C. W., *Probability and Stochastic Processes for Engineers*, 2nd edition. New York: Macmillan, 1991.

Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 3rd edition. New York: McGraw-Hill, 1991.

Scheaffer, R. L., *Introduction to Probability and Its Applications.* Belmont, CA: Duxbury, 1990.

Viniotis, Y., *Probability and Random Processes for Electrical Engineers.* New York: McGraw-Hill, 1998.

Yates, R. D. and D. J. Goodman, *Probability and Stochastic Processes*, 2nd edition, New York: Wiley, 2004.

第6章 参量估计

6.1 引言

统计学，也称为概率应用，是一门将概率原理应用于实际数据的学科。统计学的两个主要领域是参数估计和假设检验。在参数估计中，我们用实际数据来估计参数的均值，标准差，方差，协方差，概率或分布等。在假设检验中，如果可行的话，我们在一个概率背景下用实际数据来实现合理的判决，假设检验将在第7章讨论。

我们知道，概率论是一门基于公理与定义的数学理论，其主要结论是定理、推论、关系和模型。尽管我们可以应用概率论对相当广泛的问题进行建模和求解，但求解这些问题的先决条件在现实中并不是轻松获得的。举例来说，假定我们给定 $X{:}N(\mu,\ \sigma^2)$，要求计算事件 $E=\{-1\leqslant X\leqslant +1\}$ 发生的概率，可以很容易地得到结果 $F_{\rm SN}((1-\mu)/\sigma)-F_{\rm SN}((-1-\mu)/\sigma)$。然而在现实中，如何来确定参数 μ、σ？更进一步，我们如何能确定是高斯分布呢？在前面的章节中，运用几个重要的关于随机变量 X 的参数，如均值或期望 μ，标准差 σ_X，方差 σ_X^2，以及随机变量 X 和 Y 的相关性参数 $E[XY]$，等等。在实际中，用估计器来估计这些量，这些估计器是随机变量的函数。一个好的估计器的特征是什么？对于同一参数如何在不同的估计器中进行选择？如何描述一个估计量与真实但是未知值的趋近程度。

已经有很多关于参量估计问题的文章，但关于这个问题的研究并没有停止，这从档案文献中可以得到证明。有几本关于统计学和参量估计方面的好书，都带有一点工程味道（如参考文献[6-1, 6-2]），另外，在互联网上有海量的素材。

例 6.1-1 （抛硬币公平吗？）假定你加入一个抛硬币的游戏并想了解这是否公平。你抛一个硬币结果是“正面”(H)，能得到什么结论？其实除了能得到抛硬币可以抛得到“正面”的结论，很难得到其他结论。现在重复试验并观测到“背面”(T)，能得到抛硬币是公平的结论吗？显然这是一个不可靠的结论。假定抛硬币10次并观测得到{H, T, T, H, H, T, H, T, H, T}，基于观测结果并运用概率的频率解释，我们可以得到 $P[\mathrm{H}]=n_H/n=5/10=0.5$，由此得到抛硬币是公平的结论，但这结论还不是很确定。另外，如果观测得到结果是{H, T, H, H, H, H, T, H, H, H}，你会得到抛硬币倾向于得到“正面”的结论，同样也不是很确定的。那么有定量的方法来描述这种不确定性（或确定性）吗？下面，将介绍一些方法来回答这个问题。

独立同分布观测量

在上面的抛硬币试验中，可以定义一个随机变量

$$X \triangleq \begin{cases}1, & \text{H 朝上}\\ 0, & \text{T 朝上}\end{cases}$$

重复试验 n 次，得到一个随机变量序列 X_i，$i=1,\ \cdots,\ n$，我们称之为独立同分布(i. i. d)观测量。这些独立同分布观测量 $\{X_i;\ i=1,\ \cdots,\ n\}$ 称为 X 的 n 维随机样本。在某些情况下，X 被

称为全体集合，X 的观测量称为 n 维随机样本。在这个试验中，X_i是伯努利分布变量，可以推广到任意分布，只要其有相同的分布函数、概率密度或概率集合函数，并且各个观测不受前面的分布结果影响。

我们在讨论中心极限定理时已经介绍过独立同分布概念，由于其在统计学中非常重要，这里将更详细地加以阐述。在上述例子中，由于抛硬币过程不受以前或以后抛掷过程的影响，因此观测量是独立的。更严谨地，X_i，$i=1$，…，n 的联合概率集合函数

$$P_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)=P_{X_1}(x_1)P_{X_2}(x_2)\cdots P_{X_n}(x_n)$$

它们是独立分布，也由于假定在所有抛掷过程中应用同一硬币并且其不受试验影响。即有

$$P_{X_1}(x)=P_{X_2}(x)=\cdots=P_{X_n}(x)\triangleq P_X(x),\ \ -\infty<x<\infty$$

对于连续型随机变量而言独立同分布意味着

$$f_{X_1X_2\cdots X_n}(x_1,x_2,\cdots,x_n)=f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$$
$$f_{X_1}(x)=f_{X_2}(x)=\cdots=f_{X_n}(x)\triangleq f_X(x),-\infty<x<\infty$$

独立同分布的观点与很多读者的认识有冲突，比如说，对于抛硬币这样一个经过物理和统计检验的公平判决，若前 9 次抛掷结果是“正面”，一些读者会期望第 10 次结果为“背面”，以达到某种平衡。但硬币对以往的结果没有记忆，因此第 10 次“正面”“背面”的概率依然一样[①]。

例 6.1-2 （同分布条件的失去）研究一个理发店顾客到达速率。我们将一个工作日（早晨 7 点至下午 3 点）分成 16 个半小时的间隔，计算各间隔顾客数目。假定 X_i，$i=1$，…，16 表示在第 i 个间隔顾客到达数目。这里 X_i不是独立同分布的（同分布前提不满足）。我们可以期望在早晨上班前比其他时刻有更多的顾客，除了午休时段。

例 6.1-3 （有偏随机样本）一个生产荞麦食品的早餐供应公司意图说明吃荞麦食品可以减肥。为此目的，公司雇用一调查员来调查试图通过吃荞麦食品来减肥的人。调查员从吃荞麦食品的人中随机选取，但结果并不显示荞麦食品可以减肥，调查员将调查范围缩小为体重低于平均值的人。令 X_i表示第 i 个调查对象吃了荞麦食品三个月后减轻的体重，可以看到从公平调查得到的集合$\{X_i\}$与通过有偏调查得到的集合$\{X_i\}$分布是不同的。另外，我们可以将此问题描述成一个假设检验问题，即吃荞麦食品可以减肥以及不能减肥两个假设。

例 6.1-4 （非独立序列）一个保守的赌徒玩了 n 轮二十一点，他从筹码处换了 100 美元的筹码，但每轮仅下注 1 美元。令 X_i表示第 i 轮筹码的金额，则 X_i，$i=1$，…，n，是独立的序列吗？很明显，$X_{i+1}=X_i\pm1$，因此 X_i并不是相互独立的；举例来说，尽管分开来的概率都不为零，我们有 $P[X_i=10, X_{i+1}=12]=0$。令 Y_i表示赌徒在第 i 轮赢（或输）的金额，则有 $Y_i=\pm1$，那么 Y_i，$i=1$，…，n 是独立序列吗？答案是独立序列[②]，因为第 i 轮的输赢与之前或之后的结果无关。

例 6.1-5 （回顾联合概率及总和概率）假定获取三个独立同分布的 0/1 的伯努利随机变量 X，表示为 X_1，X_2，X_3[③]。X 的概率集合函数 $P_X(x)=p^xq^{1-x}$，$p+q=1$，$x=0$，1。观测量的

---

① 尽管如此，如果在大量的抛掷结果中，“正面”出现的次数明显超过“背面”出现次数，则抛硬币的公平性的假定需要重新确认。可以用假设检验（参见第 7 章）将此问题假设为“可能公平”或“可能不公平”。

② 此游戏有几个假定，一是发牌人是公正的，二是在赢或输之后不改变战略。

③ 注意此处 X_i为离散随机变量。

联合概率集合函数为

$$P_{X_1X_2X_3}(x_1,x_2,x_3)=p^{x_1+x_2+x_3}q^{3-(x_1+x_2+x_3)}$$

注意到这与和式 $Y\triangleq X_1+X_2+X_3$ 的概率集合函数是不同的，和式的概率集合函数为 $P_Y(k)=b(k;3,p)=\binom{3}{k}p^kq^{3-k}$。

概率估计

假定基于观测量的随机事件 E 的概率估计①为 $\hat{P}[E]=n_E/n=0.44$，n 为样本数，n_E 为随机事件 E 发生次数。0.44 与事件发生的真实概率的接近程度如何？而事件发生的真实概率通常是难以获得的。假定一个医学研究者想要获得其病人红血球细胞小于一般尺寸的比例(概率×100)，其真实比例应该由病人身体中所有小于一般尺寸的红血球细胞除以所有的红血球细胞获得，然而不会这么做。尽管如此，我们可以通过几滴血中获得相当接近的估计。另一个例子，假定美国的一个州有 343 065 名有效选举人，并有 144 087 投票给了共和党，则随机选出一个人，投票给共和党的概率为 0.42。然而，调查所有 343 065 名选举人的费用太高(或时间不允许)，因此，调查员需要从更小的随机样本中做出预测。我们假定调查员随机选取了 512 名选举人，其中 225 名投票给了共和党，则共和党的投票率估计为 0.44。注意到如果随机样本数足够小，则共和党投票率的估计可以为零至一中间的任何数。举例来说，如果仅调查两名选举人，而他们都投票给了共和党，则共和党的投票率估计为 1！但这个估计是完全不真实的！另一方面，如果可以说出"以接近 0.98 的置信概率估计得到共和党的投票率介于 0.42 ~ 0.44 之间"，则这个共和党投票率调查的声明就比较可靠。0.98 的概率表明我们基本可以确信共和党的投票率介于 42% ~44% 之间。而常规的投票率估计并不是一个相当确信的数据，随着调查样本的减小会变得更加不确信。在现实中，我们倾向于提供具有可靠性分析的估计结果，而不是无可靠性分析的估计结果。

参量估计的中心任务是重构必然会(或几乎会)发生的事件，也就是说，这些事件是以固定概率发生的。这并不意味着对概率 $\hat{p}$ 的估计是不可信的或无意义的。举例来说，假定一个有数千辆小车的租车公司，发现在第一年底其 n_1 辆车中有 n_E 辆由于损耗需要更换。在所有条件都不变的情况下，若该公司第二年有 n_2 辆车，则可以期望大约有 $\hat{p}n_2$ 辆车在年底会需要更换，或许会多一点，或许会少一点，其中 $\hat{p}\triangleq n_E/n_1$ 是第二年小车需要更换的概率。从公司的管理层观点看，估计值 $\hat{p}n_2$ 对第二年的计划和预算是非常有用的数据。

下述的例 6.1-6 说明了应用统计学的基本原理能够得到非常确定性的结论。

例 6.1-6 (估计湖中鱼的数目)下述问题是为了阐述统计学能够用于得到有意义的必然事件。美国渔业和野生动物协会(FWS)是一个内务机构，对巴斯湖这一大湖中鲈鱼的比例感兴趣。为此目的，会用渔网捕获一定数量的鱼来分析其组成。用 n 表示捕获鱼的数目，n_B 表示鲈鱼的数目，则鲈鱼的比例估计为 $\hat{p}=n_B/n$，由于 n_B 是随机的，因此估计也是随机量②。我们不将 n 视为随机量，因为我们可以预先设置 n 的大小。鲈鱼分布的真实概率 p 是湖中鲈鱼的数目除以鱼的总数，这个概率是未知的(由于鱼可能互相吃掉对方而为随时间稍微变化的变

---

① 在第 1 章中提到，绝大多数实际问题中，概率是需要估计的。

② 此处和其他一些地方我们没有用大写字母来表示随机变量。

量)。我们必须仔细区分随机变量n_B(是一个函数)和它的实现,将它的实现加一个上标。例如,n_B的一个实现为$n_B'=58$,假定$n=133$,则鲈鱼的概率估计为$\hat{p}'=58/133=0.44$①。函数n_B的取值范围为$[0, n]$中的整数值,而实现$\hat{p}'$是鲈鱼真实概率p的一次估计,下面我们将对巴斯湖中鲈鱼数目做一更明确的说明。假定对所有鱼样本进行一一检验,令

$$X_i \triangleq \begin{cases} 1, & 第 i 条鱼为鲈鱼 \\ 0, & 其他 \end{cases}$$

则X_i为伯努利随机变量,概率集合函数$P_{X_i}(x)=p^x(1-p)^{1-x}(x=0, 1)$。随机样本$\{X_i, i=1, \cdots, n\}$由$n$个独立同分布随机变量$X$的观测量组成,$X$即是否为鲈鱼的随机变量表示。我们可以将$X$看成鱼的种群描述。随机变量$Z \triangleq \sum_{i=1}^{n} X_i$表示$n$个样本中鲈鱼的总数,$Z/n \triangleq \hat{p}$即为$p$的估计。$Z$是独立伯努利随机变量之和,为概率集合函数$b(k; n, p)$(参见例4.8-1)的二项式随机变量,可知其均值为np,标准差为$\sigma_Y=\sqrt{np(1-p)}$。我们下面生成(几乎)确定事件$E=\{np-3\sqrt{np(1-p)} \leqslant Z \leqslant np+3\sqrt{np(1-p)}\}$。由于$Z$是独立同分布随机变量之和,根据中心极限定理,可以应用正态分布来近似计算$P[E]$。这在例4.8-3中已有计算,结果是0.997。我们也可以应用$Z/n \triangleq \hat{p}$重写$P[E]$为$P[E]=P[(p-\hat{p})^2 \leqslant \frac{9}{n}p(1-p)]=0.997$,建议读者自己验证此结果。其中参量$p$是二次的,是求解$(p-\hat{p})^2=(9/n)p(1-p)$方程的两个根$p_1, p_2$,给出的是以概率0.997生成事件对应$\hat{p}$的积分区间的两个端点值。端点值为

$$p_1, p_2 = \frac{2\hat{p}+(9/n)}{2[1+(9/n)]} \mp \sqrt{\left(\frac{2\hat{p}+(9/n)}{2[1+(9/n)]}\right)^2 - \frac{\hat{p}^2}{1+(9/n)}} \tag{6.1-1}$$

对于$n=133$, $n_B=58$,我们有$\hat{p}'\approx 0.44$,并因此得到$\hat{p}_1'\approx 0.31$, $\hat{p}_2'\approx 0.57$。怎么解释这个结果?首先注意到间隔$[0.31, 0.57]$并没有与概率相联系,这个间隔可能包含了真实概率p,也可能没有,其长度是0.26且$|\hat{p}'-\hat{p}_1'|=|\hat{p}'-\hat{p}_2'|\approx 0.13$。数值$100\times|\hat{p}'-\hat{p}_1'|$有时又称为误差范围,在此处为13%②。由于两个端点为随机变量,因此端点间隔$[p_1, p_2]$为随机间隔,也就是说,它们依赖于估值$\hat{p}$。尽管如此,平均来说,这个间隔应该在1000次实验中包含有997次p值。我们注意到湖中鲈鱼比例离0和100%很远,但我们可以确信在巴斯湖中鲈鱼比例在31% ~57%之间。

上述例子说明了统计学可以帮我们对巴斯湖的鲈鱼数量给出一个相当明确的声明,声明可以这样写:研究表明,巴斯湖中44%的鱼为鲈鱼,误差范围为±13%。

例6.1-7 (估计登革热发烧概率)报刊文章报道③,在佛罗里达州的基韦斯特岛,居民和游客受到登革热病毒威胁,这种疾病由携带病毒的蚊虫叮咬所致。尽管岛内旅游部门说游客受病毒感染的可能性不大,一项独立调查表明,在随机选取的240名居民和游客中,有13名的登革热病毒显阳性。有些人质疑样本数量太小,结果不准确,实际发热概率应该小得多。下面根据以上述随机样本数,以95%的置信度计算登革热病毒感染的真实概率的置信区间。

解 真实概率的估计为$\hat{p}=K/n$,其中K为二项式随机变量,即

---

① 估计器的实现有时称为一个估计,也就是一个数值。

② 通常用代数符号表示误差范围,例如此处用±13%。

③ 《纽约时报》2010年7月23日。

$$P_K[\text{在}\, n\, \text{次尝试中有}\, k\, \text{次成功}] \triangleq b(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

其中 $E[\hat{p}]=p$，$\text{Var}[\hat{p}]=p(1-p)/n$。根据数据我们计算得到均值估计为 $\hat{p}'=13/240=0.054$。由于 $n \gg 1$，应用二项式分布的正态近似并定义标准正态随机变量 $X:N(0,1)$ 为

$$X \triangleq (\hat{p}-p)/\sqrt{p(1-p)/n}$$

则概率 p 的95%的置信区间为求解 $P(-x_{0.975}<X<x_{0.975})=2F_{\text{SN}}(x_{0.975})-1=0.95$，即 $x_{0.975}\approx 1.96$。则

$$P[-1.96 < \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} < 1.96] = 0.95$$

应用例6.1-6中的方法，可以得到 p 的95%置信区间为多项式 $1.016p^2-0.124p+0.003=0$ 的根，即 $p_l=0.033$，$p_u=0.089$。这样，就以95%相信感染率为每30人感染1人至每11人感染1人之间。这个结果会影响到你的基韦斯特岛旅行计划吗?

6.2 估计器

估计器是随机变量的函数，用于估计参数但并不依赖于参数。下面通过几个例子来说明。

例6.2-1 （包装的真相）一个消费者保护协会（CPA）想证实超市出售的熟火鸡的鸡胸肉在包装上标注的“70%肉，30%水”的真实性。火鸡肉由“Sundry农场”制作，CPA购买了5包来检测肉的百分含量（mcp）。用 X_i 表示第 i 包的mcp，CPA采用公式 $\hat{\Theta}_1=(1/n)\sum_{i=1}^{n}X_i$ 估计平均mcp。测量得到5包火鸡肉的mcp分别为：68%，82%，71%，65%，67%，由此平均mcp为70.6%。

70.6%表示估计器 $\hat{\Theta}_1$ 的一个实现，通常称为一个估计。若CPA从“Sundry农场”购买另外5包熟的火鸡肉，计算的结果无疑会与前面得到有些细微差异。

例6.2-2 （包装的真相继续）CPA试图估计“Sundry农场”火鸡肉的肉含量变化率，采用公式 $\hat{\Theta}_2=((1/n)\sum_{i=1}^{n}(X_i-(1/n)\sum_{j=1}^{n}X_j)^2)^{1/2}$，代入 $n=5$ 并利用上题数据得到近似6%的变化率。

例6.2-3 （包装的真相继续）CPA由于在上一问题中采用 $\hat{\Theta}_2$ 来测量变化率受到质疑，建议用 $\hat{\Theta}_3=((1/(n-1))\sum_{i=1}^{n}(X_i-(1/n)\sum_{j=1}^{n}X_j)^2)^{1/2}$ 来替代。当采用上式 $\hat{\Theta}_3$ 并代入 $n=5$ 得到肉含量的变化率为6.7%。

下面我们将看到 $\hat{\Theta}_1$ 是均值的无偏、一致估计器，$\hat{\Theta}_2$ 是标准偏差的有偏、最大似然估计器，$\hat{\Theta}_3$ 是标准偏差的无偏、一致估计器。还有一些估计器用来估计 $\text{Var}[X]$、方差阵 $\boldsymbol{K}$ 和其他高阶矩。

与其他估计器相比，一些估计器会有更多值得称道的性能。为评价估计器我们先介绍下面一些定义。

定义6.2-1 估计 θ 的估计器[①] $\hat{\Theta}$ 是观测向量 $\boldsymbol{X}=[X_1,\cdots,X_n]^{\text{T}}$ 的函数，但不是 θ 的函数。

① 基于重复观测的参量估计或其他估计，如瞬时概率，理论上是基于大数定理和切比雪夫不等式。

定义 6.2-2 估计 θ 的估计器 $\hat{\Theta}$ 当且仅当 $E[\hat{\Theta}]=\theta$ 时称为无偏。估计的偏差①定义为 $E[\hat{\Theta}]=\theta$。

定义 6.2-3 如果估计 θ 的估计器 $\hat{\Theta}$ 是观测向量 $\boldsymbol{X}\triangleq[X_1, \cdots, X_n]^{\mathrm{T}}$ 的线性函数，即满足

$$\hat{\Theta}=\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X} \tag{6.2-1}$$

则 $\hat{\Theta}$ 称为线性估计器。向量 $\boldsymbol{b}$ 为 $n\times 1$ 维系数向量且不依赖于 $\boldsymbol{X}$。

定义 6.2-4 令 $\hat{\Theta}_n$ 表示基于任意 $n(n>1)$ 个样本数据 $X_1, \cdots, X_n$ 的估计器，则如果对于任意 $\varepsilon>0$，满足

$$\lim_{n\to\infty} P[|\hat{\Theta}_n-\theta|>\varepsilon]=0, \qquad \varepsilon>0 \tag{6.2-2}$$

则 $\hat{\Theta}_n$ 称为一致估计器。式(6.2.2)的条件也称为依概率收敛。

定义 6.2-5 若估计器 $\hat{\Theta}$ 满足

$$E[(\hat{\Theta}-\theta)^2]\leqslant E[(\hat{\Theta}'-\theta)^2] \tag{6.2-3}$$

则 $\hat{\Theta}$ 称为最小方差无偏估计器。其中 $\hat{\Theta}'$ 是任意其他估计器且满足 $E[\hat{\Theta}']=E[\hat{\Theta}]=\theta$。

定义 6.2-6 若估计器 $\hat{\Theta}$ 满足

$$E[(\hat{\Theta}-\theta)^2]\leqslant E[(\hat{\Theta}'-\theta)^2] \tag{6.2-4}$$

则 $\hat{\Theta}$ 称为最小均方误差(MMSE)估计器。其中 $\hat{\Theta}'$ 是任意其他估计器。

其他一些关于估计器的良好性质，如有效性、完备性和不变性。这些性质在统计学②教材里讨论，这里不做深入分析。

6.3 均值估计

在第 4 章中，我们提及基于一组数据集合的数值平均 μ_s 是与所有数据 $x_1, x_2, \cdots, x_n$ 的同时最接近的一个数，从这个意义上看，μ_s 可以认为是这个数据集合的最佳代表。用力学观点看，均值就是集合的重心。尽管均值并不能代表集合的全部信息，但其在任何情况下都是一个有用的描述。例如，在一场所有条件相同的标准化考试中，A 校学生的平均分为 92，B 校学生的平均分为 71，可以得出结论，A 校学生比 B 校学生准备得更好。若根据大量资料分析，并考虑一些相关因素(例如性别、收入、种族、生活方式等)，表明吸烟者的平均寿命为 67 岁，而不吸烟者的平均寿命为 78 岁，则可以得出吸烟有害健康的结论。

将式(4.1-1)做一些记号上的调整如下：

$$\mu_s(n)=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{6.3-1}$$

我们看到，数值平均与数据本身及数据量 n 相关。模型中我们假定数据为随机向量 X 的 n 个独立同分布实现，也就是说，x_1 是观测 X_1 的一次实现，x_2 是观测 X_2 的一次实现，以此类推。每

① 偏差定义通常是无符号的，也就是说可以为正也可以为负。

② 参见参考文献[6-1]的第 8 章。

个随机变量 X_i 都是函数而 x_i 是该函数获取的一个数值。从随机样本 $\{X_1, \cdots, X_n\}$ 中估计未知参量 $\mu_X \triangleq E[X]$，得到均值估计函数

$$\hat{\mu}_X(n) \triangleq \frac{1}{n}\sum_{i=1}^{n} X_i \tag{6.3-2}$$

我们发现 $\hat{\mu}_X(n)$ 就是在 6.2 节中的估计量 $\hat{\Theta}$。式(6.3-2)的结果称为样本均值。用上三角表示 $\hat{\mu}_X(n)$ 是一个估计量而不是实际的均值。另外，也可以得到方差估计量函数(VEF)或样本方差为

$$\hat{\sigma}_X^2(n) \triangleq \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2 \tag{6.3-3}$$

我们发现 VEF 即 6.2 节中估计量 $\hat{\Theta}_3$ 的平方。上式为两个 VEF 计算方式中的一种，另外一种为

$$\hat{\sigma}_X^2(n) \triangleq \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2 \tag{6.3-4}$$

我们将在后面的章节中讨论方差估计量。现在，要求读者采用式(6.3-3)计算方差。下面将会发现，采用 VEF 计算式(6.3-3)来估计 σ_X^2 是完全合理的，其一个合理的性质就是 $E[\hat{\sigma}_X^2] = \sigma_X^2$，而对采用式(6.3-4)来计算 VEF 是渐进正确的。

均值估计函数的性质(MEF)

式(6.3-2)计算得到的均值估计量是无偏的，即 $E[\hat{\mu}_X(n) - \mu_X] = 0$，很容易证明这个重要的结论。我们有

$$E[\hat{\mu}_X(n)] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{\mu_X}{n}\sum_{i=1}^{n} 1 = \frac{\mu_X}{n} n = \mu_X \tag{6.3-5}$$

虽然不是必需的①，但无偏总是希望有的一个性质。另外的一个重要性质是，我们希望估计量随着观测量的增加而变得更好。举例来说，希望由式(6.3-2)计算的 MEF，其基于 100 个观测量的计算结果比基于 10 个观测量的计算结果更可信。一个衡量是否可信的方法是分析无偏估计量的方差，若无偏估计量的方差很小，可以说明 $\hat{\mu}_X(n)$ 的一个实现离真实均值 μ_X 很近；若方差很大，则说明可能偏离真实均值很远。考虑 $\hat{\mu}_X(n)$ 的方差，即

$$\begin{aligned}
\sigma_{\hat{\mu}}^2(n) &\triangleq E\left[(\hat{\mu}_X(n) - (E[\hat{\mu}_X(n)]))^2\right] = E\left[(\hat{\mu}_X(n) - \mu_X)^2\right] \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)\right)^2\right] \\
&= E\left[\frac{1}{n^2}\sum_{i=1}^{n}(X_i - \mu_X)^2\right] + E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(X_i - \mu_X)(X_j - \mu_X)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} E[(X_i - \mu_X)^2] + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i\neq j}^{n} E\left[(X_i - \mu_X)(X_j - \mu_X)\right] \\
&= \sigma_X^2/n
\end{aligned} \tag{6.3-6}$$

① 如果估计器有其他良好性质，可以接受其小小的偏差。

在第一行中，公式右边应用了 MEF 的无偏性，第二行中，应用了 MEF 的定义并将 μ_X 先乘后除 n，第三行中，应用了和式的平方为各自的平方加上不同下标交叉量的乘积，第四行中，应用了均值估计器的线性性质，第五行中，应用了下述性质，即对于 $i \neq j$，X_i 和 X_j 相互独立且有 $E[(X_i-\mu_X)(X_j-\mu_X)]=0$。在此借助切比雪夫不等式(参见 4.4 节)应用于 $\hat{\mu}_X(n)$，则有：对于任意 $\delta>0$，满足

$$P[|\hat{\mu}_X(n)-\mu_X| \geqslant \delta] \leqslant \frac{\operatorname{Var}(\hat{\mu}_X(n))}{\delta^2} = \frac{\sigma_X^2}{n\delta^2} \xrightarrow{n\to\infty} 0 \tag{6.3-7}$$

式(6.3-6)和式(6.3-7)均是统计学中非常重要的结论。式(6.3-6)表明均值估计器的方差随着 n 的增加而降低，因此可以选择足够大的样本数而使得方差足够小。特别是，均值估计器的方差等于观测方差除以样本数目，这对于观测量 X 为有限方差时总是成立的。式(6.3-7)表明真实均值和 MEF 的偏差超过一个确定量这个事件当样本数据足够大时是不太可能发生的——不管这个确定量多么小。一个满足式(6.3-6)和式(6.3-7)的估计量称为一致性估计量。

例 6.3-1 (样本数对均值估计的影响)X 为正态分布且 $\sigma_X=3$，我们希望计算 $P[|\hat{\mu}(n)-\mu_X|\leqslant 0.1]$。为阐述样本数的影响，采用两种随机样本：小样本($n=64$)和大样本($n=3600$)。根据公式

$$\begin{aligned}
&P[-0.1<\hat{\mu}_X(n)-\mu_X<0.1]\\
&=P\left[-0.1\sqrt{n}/\sigma_X<Y<0.1\sqrt{n}/\sigma_X\right]\\
&=2\operatorname{erf}\left(\frac{0.1\sqrt{n}}{\sigma_X}\right)\\
&=2\operatorname{erf}(0.0333\sqrt{n})
\end{aligned}$$

其中 $Y \triangleq (\hat{\mu}_X-\mu_X)/(\sigma_X/\sqrt{n})$ 的分布为 $N(0,1)$。当 $n=64$ 时，$P[|\hat{\mu}_X(n)-\mu_X|\leqslant 0.1]\approx 0.2$，对此结果的解释是：对于 64 个样本数的均值估计，在 1000 次实验中，仅有 200 次的估计结果偏离真实均值 0.1 或更小。当 $n=3600$ 时，有 $P[|\hat{\mu}_X(n)-\mu_X|\leqslant 0.1]\approx 0.95$，即在 1000 次实验中，事件 $\{|\hat{\mu}_X(n)-\mu_X|\leqslant 0.1\}$ 大约发生 950 次，也即在一次实验中，事件 $\{|\hat{\mu}_X(n)-\mu_X|\leqslant 0.1\}$ 几乎确定会发生。

例 6.3-2 (获得均值的 95% 置信区间需要多少样本?)我们希望计算正态随机变量 X 均值的 95% 置信区间，需要多少 X 的观测量 $X_1,\cdots,X_n$ 呢？更进一步，什么参数决定了区间的长度和位置呢？术语"95% 置信区间"意味着我们在实线上寻找最小(或接近最小)区间的端点，在 1000 次实验中，会有不少于 950 次离真实均值的范围不超过这个区间。用概率表示为

$$P[|\hat{\mu}_X(n)-\mu_X| \leqslant \gamma_{0.95}]=0.95 \tag{6.3-8}$$

其中 $\gamma_{0.95}$ 是需要确定的数值，下标表示这是我们寻求的 95% 置信区间。回想 $\hat{\mu}_X(n)$ 分布为 $N(\mu_X,\sigma_X^2/n)$，因此

$$Y \triangleq \frac{\hat{\mu}_X(n)-\mu_X}{\sigma_X/\sqrt{n}} \tag{6.3-9}$$

分布为 $N(0,1)$。将 Y 代入式(6.3-8)并重写如下：

$$\begin{aligned}
0.95 &= P[-\gamma_{0.95} \leqslant \hat{\mu}_X(n)-\mu_X \leqslant \gamma_{0.95}]\\
&= P\left[-\gamma_{0.95}\sqrt{n}/\sigma_X \leqslant Y \leqslant \gamma_{0.95}\sqrt{n}/\sigma_X\right]\\
&= 2F_{\mathrm{SN}}(\gamma_{0.95}\sqrt{n}/\sigma_X)-1
\end{aligned} \tag{6.3-10}$$

上式第二行中，将随机变量 $\hat{\mu}_X(n)-\mu_X$ 变换为 $N(0,1)$分布的随机变量 Y。第三行中，将概率表

示为标准正态(SN)CDF。式(6.3-10)的最后一行就是我们寻求的结果，即有 $F_{SN}(\gamma_{0.95}\sqrt{n}/\sigma_X)=0.975$。有时候我们用符号 $F_{SN}(z_u)=u$ 表示标准正态 CDF，数值 z_u 称为标准正态的 u 百分数。从 CDF 的表格中(参见附录 G)可以得到 $z_{0.975}=1.96$，由 $z_{0.975}=\gamma_{0.95}\sqrt{n}/\sigma_X$ 可以推导 $\gamma_{0.95}\sqrt{n}/\sigma_X=1.96$，即 $\gamma_{0.95}=1.96\sigma_X/\sqrt{n}$。回到问题上来，我们注意到事件 $\{|\hat{\mu}_X(n)-\mu_X|\leqslant\gamma_{0.95}\}$ 与事件 $\{\hat{\mu}_X(n)-\gamma_{0.95}\leqslant\mu_X\leqslant+\hat{\mu}_X(n)+\gamma_{0.95}\}$ 是相同的，则从式(6.3-10)的中间一行得出 μ_X 的最小的95%置信区间为

$$\left[-1.96\frac{\sigma_X}{\sqrt{n}}+\hat{\mu}_X(n),\ 1.96\frac{\sigma_X}{\sqrt{n}}+\hat{\mu}_X(n)\right] \tag{6.3-11}$$

显然，这个结果可以推广到其他置信区间分析上。假定查找 δ 置信区间(此处 $\delta=0.95$)，则 μ_X 的 δ 置信区间为

$$\left[-z_{(1+\delta)/2}\frac{\sigma_X}{\sqrt{n}}+\hat{\mu}_X(n), z_{(1+\delta)/2}\frac{\sigma_X}{\sqrt{n}}+\hat{\mu}_X(n)\right] \tag{6.3-12}$$

怎么确定这是最小区间？这是由于正态分布的对称性，最大的概率在其中心位置。对于其他分布95%置信区间会需要更多的支持，也就是说，需要更大的区间。

回到我们的问题，若 σ_X 未知，则需要多少样本来确定最小的95%的置信区间。很明显，若选择一个足够大的区间，如在 $\hat{\mu}_X(n)$ 两侧加10倍标准差宽度，则尽管样本数很小，也一定会获得95%(或更多)的置信。但在两侧加10倍标准差宽度得到的不是最小置信区间，这个区间太大而没什么用处。因此还是假定最小置信区间，即以 $\hat{\mu}_X(n)$ 为中心，区间宽度为 $W_{0.95}=2\times1.96\times\sigma_X/\sqrt{n}$。显然，比值 $\sigma_X/\sqrt{n}$ 确定了区间的宽度。

若 σ_X 已知(实际中较少出现)，则我们可以确定需要多少样本来获得一个指定宽度置信区间。对于任意 δ 置信区间，置信区间的宽度为

$$W_\delta=2\times z_{(1+\delta)/2}\times\sigma_X/\sqrt{n} \tag{6.3-13}$$

我们从式(6.3-13)可见，当 X 的标准差增加时，置信区间将变宽(对真实均值增加了不确定性)，当样本数增加时，置信区间变窄(对真实均值降低了不确定性)。当需要的置信比增加时，置信区间会变宽，这是可以理解的。

当 σ_X 已知时的正态分布随机变量获得均值 δ 置信区间的步骤

(1)选择一个 δ 值并计算 $(1+\delta)/2$。

(2)从标准正态分布概率积累函数表中找到满足 $F_{SN}(z_{(1+\delta)/2})=(1+\delta)/2$ 的百分数 $z_{(1+\delta)/2}$。

(3)获得随机变量 X_i，$i=1$，…，n 的实现 x_i，$i=1$，…，n，计算平均值 $\mu_s=\frac{1}{n}\sum_{i=1}^{n}x_i$。

(4)计算区间 $\left[-z_{(1+\delta)/2}\frac{\sigma_X}{\sqrt{n}}+\mu_s,\ z_{(1+\delta)/2}\frac{\sigma_X}{\sqrt{n}}+\mu_s\right]$。

到目前为止，假定 σ_X 是已知的，然而，σ_X 通常是未知的(你能想到一个已知 σ_X 但 μ_X 未知的情形吗?)，这个问题的一个可行的方法是将式(6.3-11)中的 σ_X 用估计值替代，也就是式(6.3-3)的均方根 $\hat{\sigma}_X(n)$，同时假定 Y 还是正态分布。事实上由于 $\hat{\sigma}_X(n)$ 的随机性，Y 可能

不满足正态性，当样本数不是很大时结果可能会不太精确。σ_X 未知需要我们寻找其他方法来确定规定的置信区间，下面采用 t 分布的方法就是其一。

当 σ_X 未知时的正态分布随机变量的均值置信区间

一般而言，统计学中通常以代数形式出现的分布比基本概率形式要复杂一些，其中之一就是 W. S. Gossett 引入的称为“学生分布”的 t 分布，他是在分析方差未知时正态分布均值的置信区间引入的。Gossett 被认为是现代统计学奠基人之一，他更为人熟知的是他的笔名“学生”[①]。在前面的讨论中，对分布 $N(0, 1)$ 的随机变量寻求置信区间的端点

$$Y \triangleq \frac{\hat{\mu}_X(n)-\mu_X}{\sigma_X/\sqrt{n}}$$

当 σ_X 未知时，我们得不到置信区间的端点，因此，生成一个新的随机变量，其 σ_X 用下式替代

$$\hat{\sigma}_X(n)=\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2\right)^{1/2}$$

这是由式(6.3-3)的 VEF 得到的标准差值。新的随机变量定义为

$$T_{n-1} \triangleq \frac{\hat{\mu}_X(n)-\mu_X}{\left(\frac{1}{n(n-1)}\sum\limits_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2\right)^{1/2}}=\frac{\hat{\mu}_X(n)-\mu_X}{\hat{\sigma}_X(n)/\sqrt{n}} \tag{6.3-14}$$

称为 $n-1$ 自由度的 t 分布，其中 $n=2, 3, \cdots$。我们不是将 T_{n-1} 看成是标准正态随机变量的近似，而是产生一组 t 分布的随机变量。T_{n-1} 的概率密度函数用 $f_T(x;\ n-1)$ 表示，这其中的关键一点是 $f_T(x;\ n-1)$ 并没有未知参量 σ_X，从而允许我们计算均值 μ_X 的置信区间，而式(6.3-9)的随机变量 σ_X 未知则不能计算。

对于读者而言，理解产生 t 分布时没有将 σ_X 近似为 $\hat{\sigma}_X$ 是非常重要的。Gossett 的突出贡献就是采用 T_{n-1} 及其分布从而避免用正态分布来近似。

为更好理解，重写式(6.3-14)如下：

$$T_{n-1} \triangleq \frac{(\hat{\mu}_X(n)-\mu_X)\sqrt{n}/\sigma_X}{\left(\frac{1}{(n-1)}\sum\limits_{i=1}^{n}\left(\frac{X_i-\hat{\mu}_X(n)}{\sigma_X}\right)^2\right)^{1/2}}=\frac{Y}{(Z_{n-1}/n-1)^{1/2}} \tag{6.3-15}$$

其中 $Y \triangleq (\hat{\mu}_X(n)-\mu_X)\sqrt{n}/\sigma_X$: $N(0,1)$，$Z_{n-1} \triangleq \sum_{i=1}^{n}\left(\frac{X_i-\hat{\mu}_X}{\sigma_X}\right)^2$ 服从 $n-1$ 自由度的 χ^2_{n-1} 分布，其中 χ^2 称为卡方(发音为 sky-square 中的 ky-square)。卡方随机变量的下标给出的是自由度(DOF)，取值范围是(0, ∞)，意味着当 $z\leqslant 0$ 且 $n\geqslant 1$ 时，累积分布函数(CDF) $F_{\chi^2}(z;\ n)=0$。χ^2 分布在第 2 章有介绍，由于其具备 X 集合样本的独立同分布特性有时也称为样本分布。根据同样的 X_i，$i=1, \cdots, n$，虽然不是很明显，但 Y 和 Z_{n-1} 可以证明是统计独立的(参见附录 G)。根据式(6.3-14)可知，t 随机变量是标准正态随机变量(分子)与卡方随机变量除以自由度的平方根的比值。

当 n 取大值时，t 分布与正态分布相似(参见图 6.3-1)。实际上，T_{n-1} 的概率密度是以原点为中心的对称形式。

---

① 1876 - 1937，在都柏林的 Arthur Guinness & Son 啤酒厂做统计质量控制工作，由于保密原因采用笔名“学生”。



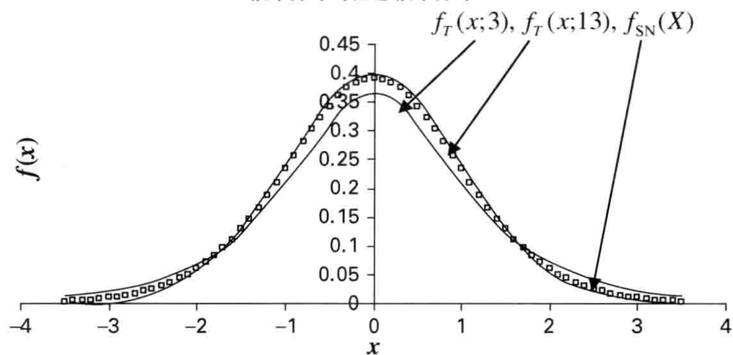


图6.3-1 T随机变量的概率密度函数与正态概率密度函数形状相似，特别当自由度较大时。图中示出当$n=3$(峰值是0.36)和$n=13$(四方块曲线，峰值是0.39)时的t概率密度函数，以及标准正态概率分布函数。除了尾部有些变动外，$n=13$时的t分布与正态分布几乎完全相同

为寻求μ_X的最小置信区间，考虑事件$\{-t_{\delta/2}\leqslant T_{n-1}\leqslant t_{\delta/2}\}$，其概率为

$$P[-t_{(1+\delta)/2}\leqslant T_{n-1}\leqslant t_{(1+\delta)/2}]=\delta \tag{6.3-16}$$

同前，μ_X的区间的置信度设定为$100\times\delta\%$。用$F_T(t;\ n-1)=\int_{-\infty}^{t}f_T(x;\ n-1)\mathrm{d}x$表示$T_{n-1}$随机变量的概率积累函数，可知

$$\delta=2F_T(t_{(1+\delta)/2},n-1)-1$$

或等价于

$$F_T(t_{(1+\delta)/2};n-1)=\frac{1+\delta}{2} \tag{6.3-17}$$

从附录G中提供的自由度$n-1$的t分布的概率积累函数表中，可以确定t百分数$t_{(1+\delta)/2}$。最后，由式(6.3-14)和式(6.3-16)，可得

$$P\left[\hat{\mu}_X(n)-\frac{t_{(1+\delta)2}\hat{\sigma}_X(n)}{\sqrt{n}}\leqslant\mu_X\leqslant\hat{\mu}_X(n)+\frac{t_{(1+\delta)/2}\hat{\sigma}_X(n)}{\sqrt{n}}\right]=\delta$$

上式给出了100δ百分比的置信区间

$$\left[\hat{\mu}_X(n)-\frac{t_{(1+\delta)/2}\hat{\sigma}_X(n)}{\sqrt{n}},\hat{\mu}_X(n)+\frac{t_{(1+\delta)/2}\hat{\sigma}_X(n)}{\sqrt{n}}\right] \tag{6.3-18}$$

置信区间的宽度为

$$W_\delta=2\frac{t_{(1+\delta)/2}\hat{\sigma}_X(n)}{\sqrt{n}} \tag{6.3-19}$$

当σ_X未知时，基于n个观测量的正态分布随机变量获得均值δ置信区间的步骤

(1)选择一个δ值并计算$(1+\delta)/2$。

(2)从T_{n-1}的概率积累函数表中找到t百分数$t_{(1+\delta)/2}$满足$F_T(t_{(1+\delta)/2};\ n-1)=(1+\delta)/2$。

(3)获得随机变量X_i，$i=1$，…，n的实现x_i，$i=1$，…，n，计算$\hat{\mu}_X(n)$，$\hat{\sigma}_X(n)$的实现。

(4)计算区间的数值实现$\left[\hat{\mu}_X(n)-\frac{t_{(1+\delta)/2}\hat{\sigma}_X(n)}{\sqrt{n}},\ \hat{\mu}_X(n)+\frac{t_{(1+\delta)/2}\hat{\sigma}_X(n)}{\sqrt{n}}\right]$。

例 6.3-3 (当 σ_X 未知时 μ_X 的置信区间 - 正态情形)正态随机变量 X 的 21 个独立同分布观测数据($n=21$)表示为 $X_1, X_2, \cdots, X_{21}$，基于观测数据分别得到 $\hat{\mu}_X(n)$ 和 $\hat{\sigma}_X(n)/\sqrt{n}$ 的实现为 3.5 和 0.45，分析 $\hat{\mu}_X(n)$ 的 90% 置信区间。

解 根据 $P[-t_{0.95} \leq T_{20} \leq t_{0.95}]=0.9$，由式(6.3-17)可得 $F_T(t_{0.95}, 20)=0.5(1+0.9)=0.95$。根据 $F=0.95$，$n=20$ 由学生分布表可得 $t_{0.95}=1.725$，由式(6.3-18)得对应的区间为 $[3.5-1.725\times0.45, 3.5+1.725\times0.45]=[2.72, 4.28]$，区间宽度为 $W_\delta \approx 2\times1.725\times0.45=1.55$。

置信区间的解释

从一组实现中得到的置信区间可能包含也可能不包含 X 的真实均值，真实均值对于我们是未知数。那么，我们说"90%"置信区间是什么意思呢？这个问题的回答涉及概率的核心意义，也就是在重复试验中想要的结果。简洁地说，"90%"置信区间意味着在 1000 次试验中，区间将包括 900 次的真实均值。是否我们将正好包括 900 次的均值呢？不一定，但 900 次的成功率是最高的。

6.4 方差和协方差估计

对于均值为 μ_X、方差为 σ_X^2 的正态随机变量 X 有 n 个观测 $X_1, X_2, \cdots, X_n$，若 μ_X 已知，则无偏 VEF 可由随机样本计算

$$\hat{\sigma}_X^2(n)=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_X)^2 \tag{6.4-1}$$

不难证明 $\hat{\sigma}_X^2$ 是 σ_X^2 的无偏的、一致估计量。若均值未知，则 VEF

$$\hat{\sigma}_X^2(n)=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2 \tag{6.4-2}$$

是 σ_X^2 的无偏的、一致估计量。

式(6.4-2)的 $\hat{\sigma}_X^2(n)$ 的无偏性 由

$$\begin{aligned} &E\left[\sum_{i=1}^{n}\left(X_i-\frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right] \\ &=E\left[\sum_{i=1}^{n}\left\{X_i^2-\frac{2}{n}X_i^2-\frac{2}{n}\sum_{\substack{j=1\\ j\neq i}}^{n}X_iX_j+\frac{1}{n^2}\sum_{j=1}^{n}X_j^2+\frac{2}{n^2}\sum_{k=1}^{n}\sum_{j>k}^{n}X_jX_k\right\}\right] \\ &=(n-1)\sigma^2 \end{aligned} \tag{6.4-3}$$

在式(6.4-3)中，我们应用了结论 $E[X_i^2]=\sigma^2+\mu^2$，$i=1, \cdots, n$。显然，若有

$$E\left[\sum_{i=1}^{n}(X_i-\hat{\mu})^2\right]=(n-1)\sigma^2$$

则有

$$E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\hat{\mu})^2\right]=\sigma^2 \tag{6.4-4}$$

又有方括号中的值就是式(6.4-2)的 $\hat{\sigma}_X^2(n)$，因此 $\hat{\sigma}_X^2(n)$ 是 σ^2 的无偏估计。

式(6.4-2)的 $\hat{\sigma}_X^2(n)$ 的一致性　$\hat{\sigma}_X^2(n)$ 的方差为

$$\begin{aligned}\mathrm{Var}[\hat{\sigma}_X^2(n)]&=E[(\hat{\sigma}_X^2(n)-\sigma^2)^2]\\&=E\left[\frac{1}{(n-1)^2}\left\{\sum_{i=1}^{n}(X_i-\hat{\mu})^4+\sum_{i\neq j}^{n}\sum^{n}(X_i-\hat{\mu})^2(X_j-\hat{\mu})^2\right\}\right.\\&\quad\left.+\sigma^4-\frac{2\sigma^2}{n-1}\sum_{i=1}^{n}(X_i-\hat{\mu})^2\right]\end{aligned}$$

直接计算可得，对于 $n\gg1$，有

$$\mathrm{Var}[\hat{\sigma}_X^2(n)]\approx\frac{1}{n}c_4 \tag{6.4-5}$$

其中 $c_4\triangleq E[(X_1-\mu)^4]$[参见式(4.3-2a)]。假定 c_4(四阶中心矩)存在，应用切比雪夫不等式，有

$$P[|\hat{\sigma}_X^2(n)-\sigma^2|>\varepsilon]\leqslant\frac{\mathrm{Var}[\hat{\sigma}_x^2(n)]}{\varepsilon^2}\approx\frac{c_4}{n\varepsilon^2}\xrightarrow{n\to\infty}0 \tag{6.4-6}$$

即 $\hat{\sigma}_X^2(n)$ 是 σ^2 的一致估计。

例 6.4-1　(正态随机变量样本均值和样本方差的数值计算)正态随机变量 X:$N(3, 1/10)$ 的 10 个观测数据为 3.12, 2.87, 3.04, 2.77, 2.89, 3.34, 3.51, 2.44, 3.28, 2.95。计算样本均值和样本方差的过程如下：

样本均值计算

$$\mu_s=\frac{1}{10}(3.12+2.87+3.04+2.77+2.89+3.34+3.51+2.44+3.28+2.95)=3.02$$

样本方差计算

$$\begin{aligned}\sigma_s^2&=\frac{1}{9}(0.01+0.225+0.0004+0.0625+0.0169+0.1024+0.2401\\&\quad+0.3364+0.0676+0.0049)\\&=0.096\end{aligned}$$

在信号处理中，比值$(\mu_s/\sigma_s)^2$ 有时称为信号噪声功率比，在此例中为 95，以分贝(dB)表示则为 $10\times\lg 95=19.8$ dB。

正态随机变量方差的置信区间

确定方差的置信区间涉及 χ^2 分布。假定正态随机变量 X 的 n 个独立同分布观测为 X_1, X_2, …, X_n，则对每一个 i

$$U_i\triangleq\frac{X_i-\mu_X}{\sigma_X} \tag{6.4-7}$$

为$N(0, 1)$分布，$Z_n \triangleq \sum_{i=1}^{n} U_i^2$是自由度为$n$的卡方分布，其概率密度函数$f_{\chi^2}(x;\ n)$如图6.4-1所示。若在式(6.4-7)中$\mu_X$未知，我们用式(6.3-2)的$\hat{\mu}_X(n)$替代，得到一个新的随机变量

$$V_i \triangleq \frac{X_i - \hat{\mu}_X(n)}{\sigma_X} \tag{6.4-8}$$

其和式$Z_{n-1} = \sum_{i=1}^{n} V_i^2$为$n-1$自由度的卡方分布。

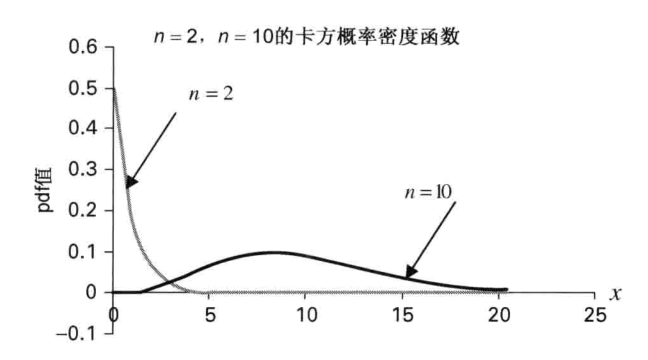


图6.4-1　$n=2$(原点值为0.5的曲线)和$n=10$的卡方概率密度函数。对所有$n>2$，概率密度函数原点值为零

例6.4-2　(卡方随机变量自由度的计算)假定V_i定义如式(6.4-8)所示，随机变量$\sum^{2} V_i^2$为1自由度的卡方分布，我们可以借助代数学来分析。可知$V_1^2+V_2^2=[(X_1-X_2)/\sigma_X\sqrt{2}]^2$，又有$U \triangleq (X_1-X_2)/\sigma_X\sqrt{2}$为$N(0, 1)$分布，因此和式$Z_n \triangleq \sum_{i=1}^{n} U_i^2$只有一个非零项，即$U^2=Z_1$。

为计算σ_X^2的δ(如$\delta=0.95$，$\delta=0.98$，$\delta=0.99$)水平的置信区间，有

$$W_{n-1} \triangleq \sum_{i=1}^{n} V_i^2 = \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2$$

寻找数值a, b使得$P[a \leqslant W_{n-1} \leqslant b] = P\left[a \leqslant \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2 \leqslant b\right] = \delta$，对于$a>0$，$b>0$且$b>a$，事件$\{\zeta: a \leqslant \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2 \leqslant b\}$等价于$\{\zeta: \frac{1}{b}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2 \leqslant \sigma_X^2 \leqslant \frac{1}{a}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2\}$，则方差的置信区间为①

$$W_\delta(a, b) = (\frac{1}{a} - \frac{1}{b})\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2 \tag{6.4-9}$$

由于$W_{n-1} = \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2$服从$\chi_{n-1}^2$分布，可由$P[a \leqslant W_{n-1} \leqslant b] = F_{\chi^2}(b;\ n-1) - F_{\chi^2}(a;\ n-1)$求解$a$, b的值。为避免计算最小区间时复杂的数学计算，我们采用获得接近最小区间的方法来计算a, b：W_{n-1}位于区间外的概率为$1-\delta=1-F_{\chi^2}(b;\ n-1)+F_{\chi^2}(a;\ n-1)$，

① 请不要混淆宽度符号$W_\delta(a, b)$和χ^2随机变量符号W_n。

若将 $1-\delta$ 描述为“错误概率”，指定 $1-F_{\chi^2}(b;\ n-1)=(1-\delta)/2$ 及 $F_{\chi^2}(a;\ n-1)=(1-\delta)/2$，则我们将“错误概率”均等地分解到 χ^2_{n-1} 概率密度函数的尾部，有 $a=x_{(1-\delta)/2}$，满足 $F_{\chi^2}(x_{(1-\delta)/2};\ n-1)=(1-\delta)/2$，$b=x_{(1+\delta)/2}$，满足 $F_{\chi^2}(x_{(1+\delta)/2};\ n-1)=(1+\delta)/2$。数值 $x_{(1-\delta)/2}$ 及 $x_{(1+\delta)/2}$ 分别称为 χ^2_{n-1} 随机变量的 $(1-\delta)/2$ 百分数和 $(1+\delta)/2$ 百分数。方差的 δ 置信区间为

$$\left\{\frac{1}{x_{(1+\delta)/2}}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2,\frac{1}{x_{(1-\delta)/2}}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2\right\}$$

其宽度 L 为

$$\left\{\left(\frac{1}{x_{(1-\delta)/2}}-\frac{1}{x_{(1+\delta)/2}}\right)\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2\right\}$$

例6.4-3 有 $X{:}N(\mu_X,\ \sigma_X^2)$ 的独立同分布观测数据16个，分析 σ_X^2 的置信区间。采用“等错误概率”方法计算 σ_X^2 接近95%的最小置信区间的 a，b 的数值。

解 $F_{\chi^2}(a;\ 15)=F_{\chi^2}(x_{0.025};\ 15)=0.025$，$F_{\chi^2}(x_{0.975};\ 15)=0.975$，从卡方分布的表中可得 $a=x_{0.025}=6.26$，$b=x_{0.925}=27.5$。

直接估计标准差

我们可估计标准差 σ_X 为

$$\hat{\sigma}_X(n)=\left(\frac{1}{n-1}\sum_{i=1}^{n}[X_i-\hat{\mu}_X(n)]^2\right)^{1/2} \tag{6.4-10}$$

这实际上需要先计算 $\hat{\sigma}_X^2(n)$，另一方法可直接估计 σ_X。考虑随机变量 X 的两个独立同分布观测 X_1：X_2，令 $Z\triangleq\max(X_1,\ X_2)$，$\hat{\mu}\triangleq(X_1+X_2)/2$，容易得到 Z 的概率密度函数为 $F_Z(z)=2F_X(z)f_X(z)$，其中 $F_X(z)$ 和 $f_X(z)$ 分别是 X 的累积分布函数和概率密度函数。考虑估计量 $\hat{\sigma}_X$

$$\hat{\sigma}_X\triangleq\sqrt{\pi}(Z-\hat{\mu}_X) \tag{6.4-11}$$

并计算 $E[\hat{\sigma}_X]\triangleq\sqrt{\pi}E[(Z-\hat{\mu}_X)]=\sqrt{\pi}(E[Z]-\hat{\mu}_X)$。当 X 为正态时，$E[Z]$ 的计算可由标准计分表[参见 *Handbook of Mathematical Functions*，M. Abramowitz，I. A. Stegun，Dover，New York，1970，p. 303，式(7.4.14)]得到，也可通过 Maple，MathCAD，Mathematica 等软件计算得到。有 $E[Z]=\mu_X+\frac{1}{\sqrt{\pi}}\sigma_X$，因此 $E[\hat{\sigma}_X]=\sigma_X$，即 $\hat{\sigma}_X\triangleq\sqrt{\pi}(Z-\hat{\mu}_X)$ 为 σ_X 的无偏估计量。

例6.4-4 （σ_X 的一次性估计）$X{:}N(\mu_X,\ \sigma_X^2)$ 的两个实现为3.8，4.1，则（上撇号表示）$Z'=\max(3.8,\ 4.1)=4.1$，$\hat{\mu}'=(3.8+4.1)/2=3.95$，$\hat{\sigma}_X'=0.26$，$\hat{\sigma}_X'$ 由式(6.3-6)可得0.21。

为计算标准差估计函数(SDEF)的方差，根据式(6.4-11)，我们有

$$\mathrm{Var}(\hat{\sigma}_X)=\pi(E[Z^2]+E[\hat{\mu}_X^2]-2E[Z\hat{\mu}_X])-\sigma_X^2$$

计算上式需要花点工夫，结果是

$$\mathrm{Var}(\hat{\sigma}_X)=\left(\frac{\pi}{2}-1\right)\sigma_X^2\approx 0.57\sigma_X^2 \tag{6.4-12}$$

实际中我们不会仅仅依据 X 的两个观测来估计 σ_X，假定有 X 的 n(偶数)个观测 $X_1, X_2, \cdots, X_n$，并将它们配对成$\{X_1, X_2\}, \cdots, \{X_{n-1}, X_n\}$，令

$$\hat{\sigma}_X^{(1)} \triangleq \sqrt{\pi}\,(\max(X_1, X_2) - 0.5(X_1 + X_2))$$
$$\hat{\sigma}_X^{(2)} \triangleq \sqrt{\pi}\,(\max(X_3, X_4) - 0.5(X_3 + X_4))$$
$$\vdots$$
$$\hat{\sigma}_X^{(n/2)} \triangleq \sqrt{\pi}\,(\max(X_{n-1}, X_n) - 0.5(X_{n-1} + X_n))$$

并定义

$$\hat{\sigma}_{\text{ave}} \triangleq \frac{1}{n/2}\sum_{i=1}^{n/2}\hat{\sigma}_X^{(i)} \tag{6.4-13}$$

则 $\text{Var}(\hat{\sigma}_{\text{ave}}) \triangleq \frac{4}{n^2}\sum_{i=1}^{n/2}\text{Var}(\hat{\sigma}_X)$，有

$$\text{Var}(\hat{\sigma}_{\text{ave}}) \approx \frac{1.04}{n}\sigma_X^2 \tag{6.4-14}$$

这明确说明了 $\hat{\sigma}_{\text{ave}}$是 σ_X 的一致估计量，我们将这留给读者作为练习。采用 $\hat{\sigma}_X \triangleq \sqrt{\pi}(Z - \hat{\mu})$来估计 σ_X 的 σ_X 置信区间将在参考文献[6-3]和参考文献[6-4]讨论。

协方差估计

协方差，定义为

$$c_{11} \triangleq \text{Cov}[XY] = E\left[(X - \mu_X)(Y - \mu_Y)\right] \tag{6.4-15}$$

其经典估计是协方差估计函数(CEF)

$$\hat{c}_{11} \triangleq \frac{1}{n-1}\sum\nolimits_{i=1}^{n}(X_i - \hat{\mu}_X(n)) \times (Y_i - \hat{\mu}_Y(n)) \tag{6.4-16}$$

其中$\{X_i, Y_i, i = 1, \cdots, n\}$为 n 对独立同分布观测，我们将留给读者来证明 $\hat{c}_{11}$是 c_{11}的无偏一致估计量。归一化协方差，也称为相关系数，定义为

$$\rho_{XY} \triangleq \frac{c_{11}}{\sqrt{\sigma_X^2\sigma_Y^2}} \tag{6.4-17}$$

可用下式估计

$$\hat{\rho}_{XY} \triangleq \frac{\hat{c}_{11}}{\sqrt{\hat{\sigma}_X^2\hat{\sigma}_Y^2}} = \frac{\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n)) \times (Y_i - \hat{\mu}_Y(n))}{\left(\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2\sum_{i=1}^{n}(Y_i - \hat{\mu}_Y(n))^2\right)^{1/2}} \tag{6.4-18}$$

$\hat{\rho}_{XY}$分布的闭合形式难以得到，但 ρ_{XY}的置信区间可用更加先进的方法得到[6-1]。

6.5 均值和方差的同时估计

若同时寻求 μ_X 和 σ_X^2 的 95% 置信区间，我们将利用随机变量 $\hat{\mu}_X(n)$ 和 $\hat{\sigma}_X^2(n)$ 的独立性，这样，我们有

$$P\left[-a \leqslant \frac{\hat{\mu}_X(n) - \mu_X}{\sigma_X/\sqrt{n}} \leqslant a, b \leqslant \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i - \hat{\mu}_X(n))^2 \leqslant c\right] = 0.95 \tag{6.5-1}$$

或等效

$$P\left[-a \leqslant \frac{\hat{\mu}_X(n)-\mu_X}{\sigma_X/\sqrt{n}} \leqslant a\right] \times P\left[b \leqslant \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2 \leqslant c\right] = 0.95 \tag{6.5-2}$$

式(6.5-2)是从式(6.5-1)推导而来，利用了下述事件的独立性

$$E_1 \triangleq \left\{-a \leqslant \frac{\hat{\mu}_X(n)-\mu_X}{\sigma_X/\sqrt{n}} \leqslant a\right\} \quad \text{和} \quad E_2 \triangleq \left\{b \leqslant \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2 \leqslant c\right\}$$

我们注意到

$$Z \triangleq \frac{\hat{\mu}_X(n)-\mu_X}{\sigma_X/\sqrt{n}}$$

是标准正态随机变量$N(0,1)$，服从分布函数$F_{SN}(z)$，同时

$$W_n \triangleq \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2$$

服从χ_{n-1}^2分布，其分布函数为$F_{\chi^2}(x;n-1)$。

下一步是将事件E_1，E_2的概率联系起来。举个例子，我们将δ置信度分解为$\delta=\sqrt{\delta}\times\sqrt{\delta}$，假定$\delta=0.95$，有

$$P\left[-a \leqslant \frac{\hat{\mu}_X(n)-\mu_X}{\sigma_X/\sqrt{n}} \leqslant a\right] = \sqrt{0.95} \approx 0.975 \tag{6.5-3}$$

和

$$P\left[b \leqslant \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2 \leqslant c\right] = \sqrt{0.95} \approx 0.975 \tag{6.5-4}$$

从式(6.5-3)我们得到$a=z_{0.9875}$，即$F_{SN}(z_{0.9875})=0.9875$，标准正态随机变量的98.75百分点。从式(6.5-4)，采用给卡方概率密度函数“等错误”分派原则可确定$b=x_{0.0125}$和$c=x_{0.9875}$，即累积卡方分布$F_{\chi^2}(x;n-1)$1.25百分点和98.75百分点。更一般地，对于任意给定的δ置信区间和给定的n，可以找到满足置信约束的a，b和c，进而可以从μ，σ^2的参数空间找到μ_X，σ_X^2的δ置信区间的边界。事件E_1是$\sigma^2=n(\mu-\hat{\mu}_X)^2/a^2$描述抛物线围成的凸区域，事件$E_2$是下述两端点间的区域

$$\begin{aligned}\sigma_{\text{Max}}^2 &= \tfrac{1}{b}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2 \ (\text{上边界})\\ \sigma_{\text{Min}}^2 &= \tfrac{1}{c}\sum_{i=1}^{n}(X_i-\hat{\mu}_X(n))^2 \ (\text{上边界})\end{aligned} \tag{6.5-5}$$

事件$E_1\cap E_2$即图6.5-1的阴影区域。

在1000次事例中约有950次，图6.5-1阴影区域会包含未知的均值和方差的值μ_X，σ_X^2。

例6.5-1 （均值和方差的置信区间）正态随机变量$X:N(\mu_X,\sigma_X^2)$有21个观测$\{X_i, i=1,\cdots,21\}$，试计算μ_X，σ_X^2的90%置信区间。

为实现90%的置信区间，在区间$(-a,a)$给$N(0,1)$随机变量Z分派（大约）0.95的概率，在区间(b,c)给自由度20的卡方随机变量W分派0.95的概率。从式(6.5-3)可得$P[-z_{0.975}<Z\leqslant z_{0.975}]=0.95$，从标准正态分布表中可得$F_{SN}(z_{0.975})=0.975$和$z_{0.975}=1.96$。从式(6.5-4)可得$P[b<W\leqslant c]=0.95$，采用例6.3-3中“等错误”分派方法可以确定$b=x_{0.025}$，

$c = x_{0.975}$，根据 $F_{\chi^2}(x_{0.025};20)=0.025$ 和 $F_{\chi^2}(x_{0.975};20)=0.975$，可得 $x_{0.025}=9.59$ 和 $x_{0.975}=34.2$，数值 $x_{0.025}$ 和 $x_{0.975}$ 即 χ^2 随机变量的 2.5 百分点和 97.5 百分点。

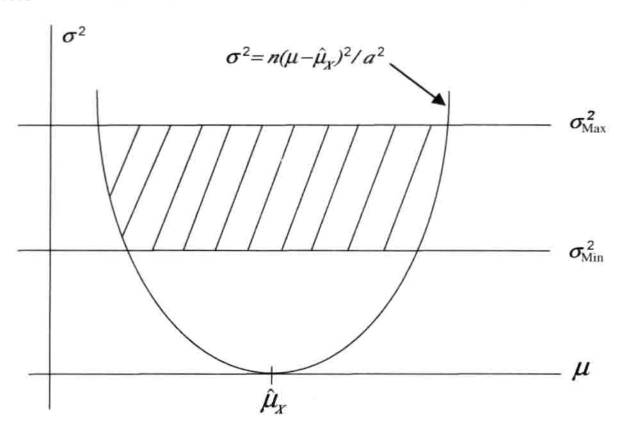


图 6.5-1　μ 和 σ^2 的联合估计的置信区间

6.6　大数据量时非高斯参量估计

考虑均值 μ、有限方差 σ^2 的随机变量 X，其有 n 个独立同分布的观测 $X\{X_i, i=1, \cdots, n\}$，根据中心极限定理可得样本均值估计(SME)①

$$\hat{\mu}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

对于大的 n 近似为正态 $N(\mu, \sigma^2/n)$。若 X 为连续随机变量，则 SME 的密度近似正态，否则其分布近似正态。当要估计的参数与非高斯分布相关时，仍然可能采用式(6.6-1)来估计置信区间

$$P\left[-a \leqslant \frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}} \leqslant a\right] = \delta \tag{6.6-1}$$

也可写成

$$P\left[(-a\sigma/\sqrt{n}) + \hat{\mu} \leqslant \mu \leqslant (a\sigma/\sqrt{n}) + \hat{\mu}\right] = \delta \tag{6.6-2}$$

读者需意识到这是 μ 的 $100\times\delta\%$ 置信区间的表达式。当分布为非高斯时，均值和方差可能是相关的，即 $\sigma=\sigma(\mu)$。如何处理这种情况？我们用参考文献[6-2]中两个例子来阐述。

例 6.6-1　(指数分布中 λ 的置信区间)假定我们要估计指数概率 $f_X(x)=\lambda e^{-\lambda x}u(x)$ 中的 λ，我们有

$$\mu \triangleq E[X] = \int_0^{\infty} x\lambda e^{-\lambda x} dx = \lambda^{-1}$$

及

$$\sigma^2 \triangleq E[(X-\mu)^2] =) \int_0^{\infty} (X-\lambda^{-1})^2 \lambda e^{-\lambda x} dx = \lambda^{-2}$$

---

① 回想可交换的应用均值估计函数和样本均值估计。

将结果代入式(6.6-2)，整理成 λ 形式

$$P\left[\frac{(-a/\sqrt{n})+1}{\hat{\mu}} \leqslant \lambda \leqslant \frac{(a/\sqrt{n})+1}{\hat{\mu}}\right]=\delta \tag{6.6-3}$$

$Z \triangleq (\hat{\mu}-\mu)\sqrt{n}/\sigma$ 近似为 $N(0, 1)$ 随机变量可获得数值 a，即满足 $F_{\mathrm{SN}}(z_{(1+\delta)/2})=(1+\delta)/2$ 可得 $a=z_{(1+\delta)/2}$，因此 λ 的百分之 $100\times\delta\%$ 的置信区间宽度为

$$W_{\delta}=2z_{(1+\delta)/2}/\hat{\mu}\sqrt{n} \tag{6.6-4}$$

例 6.6-2 （λ 置信区间的数值估计）我们要从指数随机变量 X 的 64 个独立同分布观测中获取参数 λ 的 95% 的置信区间。有 $\hat{\mu}'_X=3.5$，从式(6.6-1)可得 $2\times\mathrm{erf}(a)=0.95$，或等效 $F_{\mathrm{SN}}(z_{(1+\delta)}/2)=(1+\delta)/2=0.975$，可得 $z_{0.975}=1.96$。从式(6.6-3)和式(6.6-4)可计算出 λ 的 95% 的置信区间为{0.22, 0.36}，宽度近似为 0.14。

例 6.6-3 （伯努利分布 p 的置信区间）给定伯努利随机变量 X，概率 $P[X=1]=p$ 和 $P[X=0]=q=1-p$，我们要从 X 的 n(足够大)个独立同分布观测中以 $100\times\delta\%$ 的置信度估计 p。从分布可得 $\mu_X \triangleq E[X]=p$，MEF 为 $\hat{p}=(1/n)\sum_{i=1}^{n}X_i$。从前面章节中有 $E[\hat{p}]=p$ 及 $\mathrm{Var}[\hat{p}]=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i]=\frac{1}{n^2}npq=pq/n$，因此随机变量

$$Z \triangleq \frac{\hat{p}-p}{\sqrt{pq/n}} \tag{6.6-5}$$

在大 n 时分布为 $N(0, 1)$ 的正态(由于 X 为离散随机变量，因此不是正态密度)。为获得 p 的 $100\times\delta$ 置信区间，有

$$P[-a \leqslant \frac{\hat{p}-p}{\sqrt{pq/n}} \leqslant a]=\delta \tag{6.6-6}$$

或等效

$$P[(p-\hat{p})^2 \leqslant a^2pq/n]=\delta$$

我们从 $2\times\mathrm{erf}(a)=\delta$ 获得常数 a，即① $a=z_{\frac{1+\delta}{2}}$，通过求解 $(p-\hat{p})^2-a^2pq/n=0$(其中 $q=1-p$)的根得到置信区间的端点，即有

$$p_1=\frac{2\hat{p}+(a^2/n)}{2(1+(a^2/n))}-\frac{1}{2(1+(a^2/n))}\sqrt{(a^2/n)[(a^2/n)+4\hat{p}\hat{q}]}$$

$$p_2=\frac{2\hat{p}+(a^2/n)}{2(1+(a^2/n))}+\frac{1}{2(1+(a^2/n))}\sqrt{(a^2/n)[(a^2/n)+4\hat{p}\hat{q}]}$$

伯努利概率 p 的 95% 区间，间隔宽度为

$$W_{\gamma_a}=|p_2-p_1|=\frac{1}{(1+(a^2/n))}\sqrt{(a^2/n)[(a^2/n)+4\hat{p}\hat{q}]} \tag{6.6-7}$$

图 6.6-1 显示间隔宽度随样本数缓慢下降。

例 6.6-4 （抛硬币的公正性）我们希望获取抛硬币的公正性的一些信息。为此，抛硬币 100 次，47 次观测到"正面"，求得到"正面"的概率 p 的 95% 置信区间。应用 MEF 可得 $\hat{p}'=0.47$，从 $2\times\mathrm{erf}(a)=0.95$ 可得 $a=1.96$，进而从式(6.6-7)可得 $W_{\delta}\approx 0.192$。置信区间为

① 回想 $2\times\mathrm{erf}(a)=2\times F_{\mathrm{SN}}(a)-1=\delta$，因此 $a=x_{(1+\delta)/2}$，即标准正态随机变量的 $(1+\delta)/2$ 百分数。

0.37 ~0.57 并以 0.47 为中心。区间包含了抛硬币“公正”值 $p=0.5$，我们不能怀疑硬币是有偏的。如果独立同分布的观测增加到1200 次，其中564 为“正面”，则 $\hat{p}'$ 依然为 $\hat{p}'=0.47$，但其 95% 区间为{0.442，0.492}，不包含“公正”值 0.5，这将说明硬币有轻微有朝更多出现“背面”的方向偏离。

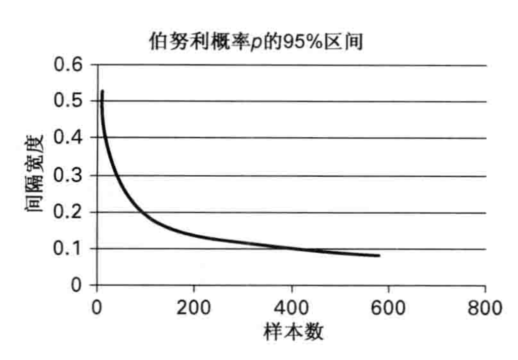


图 6.6-1　例 6.6-3 置信区间宽度随样本数的增加而缓慢下降，假定 $4\hat{p}\hat{q}\approx 1$

6.7　最大似然估计

前面我们介绍了随机变量的均值、方差、协方差估计器，尽管这些估计器具有良好的性质，它们看上去没有遵循一般准则。在本节，我们讨论通过一般准则来寻找估计器，这个方法称为最大似然(ML)准则，推导出的估计器称为最大似然估计器(MLE)。最大似然方法的缺陷是观测数据的概率密度函数的基本形式必须已知，最大似然估计方法的概念在下例中阐述。

例 6.7-1　考虑伯努利随机变量，其概率质量函数 PMF 为 $P_X(k)=p^k(1-p)^{1-k}$，其中 $P[X=1]=p$ 且 $P[X=0]=1-p$。我们将通过估计器 $\hat{p}$ 来估计 p 的值，它是 X 观测值的函数。假定有 X 的 n 个观测 $X_1, X_2, \cdots, X_n$，则 $Y=\sum_{i=1}^{n}X_i$ 即为 n 次实验中观测到 1 的次数，例如在一个已知出现“正面”的概率为 p 的投掷硬币实验，n 次的实验中出现“正面”的次数，即$\{X=1\}$。假定数值为 k_1，则出现 k_1 的先验概率为 $P[Y=k_1;p]=\binom{n}{k_1}p^{k_1}(1-p)^{n-k_1}$，由于 p 假定未知，可以看到结果是与 p 有关的。现在我们问：p 值取多少最有可能出现这种结果？由于公式右边是 p 的连续函数，可以对其微分并令等于零

$$\frac{\mathrm{d}P[Y=k_1;p]}{\mathrm{d}p}=\binom{n}{k_1}p^{k_1-1}(1-p)^{n-k_1-1}[k_1(1-p)-p(n-k_1)]=0$$

有三个根：$p=0$，$p=1$ 及 $p=k_1/n$，前两个根对应最小值，$p=k_1/n$ 对应最大值，因此，在此例中 p 的最大可能取值为 k_1/n。若将实验重复一次，有 k_2 次观测到“正面”，则 p 的估计为 k_2/n。通过最大似然估计得到 p 的估计为

$$\hat{p}=\frac{\sum_{i=1}^{n}X_i}{n} \tag{6.7-1}$$

上例中我们应用了 $\sum_{i=1}^{n}X_i$ 为二项式分布的结论，若没有此知识，还有同样的结果吗？毕竟对随机变量之和的分布的计算是要花点功夫的。回答是可以，其结果是基于生成的似然函数。

定义 6.7-1　随机变量 X_1, X_2, …, X_n 的似然函数[1]是未知参数 θ 的联合概率密度函数 $f_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n; \theta)$。特别当 X_1, X_2, …, X_n 是概率密度函数为 $F_X(x; \theta)$ 的随机变量 X 的独立观测时，则 $X_1 = x_1$, $X_2 = x_2$, …, $X_i = x_i$, …, $X_n = x_n$ 时，似然函数为

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i; \theta) \tag{6.7-2}$$

因为 $\{X_i\}$ 是概率密度函数为 $f_X(x;\theta)$ 独立同分布的随机变量。如果对于给定的结果 $X = (x_1, x_2, \cdots, x_n)$，$\theta^*(x_1, x_2, \cdots, x_n)$ 是使 $L(\theta)$ 最大的 θ 值，则 $\theta^*(x_1, x_2, \cdots, x_n)$ 就是 θ(一个数)的最大似然估计，$\hat{\theta} = \theta^*(X_1, X_2, \cdots, X_n)$ 是 θ 的最大似然估计(一个随机变量)。因此，似然函数定义为随机变量 $L(\theta) \triangleq \prod_{i=1}^{n} f_X(X_i; \theta)$ 是十分合理的，并且，使其最大化就得到 θ 的最大似然估计 $\hat{\theta}(X_1, \cdots, X_n)$。

例 6.7-2　考虑例 6.7-1 中应用似然函数求解 p 的最大似然估计。假定伯努利随机变量 X 的 n 个独立同分布观测为 X_1, X_2, …, X_n，似然函数即为 $L(\theta) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i} \times (1-p)^{n-\sum_{i=1}^{n} x_i}$。令 $\mathrm{d}L(\theta)/\mathrm{d}\theta = 0$，我们得到三个根：$p = 0$，$p = 1$ 及 $p = \sum_{i=1}^{n} x_i/n$，前两个根对应最小值，最后一个根对应最大值。因此，有 $p^*(\boldsymbol{x}) = \sum_{i=1}^{n} x_i/n$ 及 p 的最大似然估计 $\hat{p} = p^*(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} X_i/n$。

很多情况下对似然函数的对数形式求微分更方便，对数似然函数 $\lg L(\theta)$(通常为自然对数)与 $L(\theta)$ 的最大处对应同样的 θ 值。另外，微分并不能保证得到最大似然估计，这时需要寻求其他方法。最后，多参量最大似然估计可以通过求解联立方程得到，我们将通过下面三个例子分别阐述上述三点。

例 6.7-3　假定 X:$N(\mu, \sigma^2)$，其中 σ 已知，计算均值 μ 的最大似然估计。

解　X 的 n 个实现的似然函数为

$$L(\mu) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right) \tag{6.7-3}$$

由于对数形式是单调的，$L(\mu)$ 与 $\log L(\mu)$ 的最大化等效，因此

$$\log L(\mu) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

令

$$\frac{\partial \log L(\mu)}{\partial \mu} = 0$$

可得

$$\sum_{i=1}^{n}(x_i - \mu) = 0$$

因此，μ 的值，即最大化 $L(\mu)$ 的 μ^* 为

---

① 严格来说，我们应当写成 $L(\theta; x_1, x_2, \cdots, x_n)$，或如有些书上写成 $L(\theta; X_1, X_2, \cdots, X_n)$，然而，我们摒弃这类过渡标记法。

$$\mu^* = \frac{1}{n}\sum_{i=1}^{n} x_i$$

表明μ的最大似然估计应为

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{6.7-4}$$

可见，在正态情况下，μ的最大似然估计可以通过对数似然函数的微分求解，其结果是样本均值。

例 6.7-4 假定X为$(0, \theta)$间的均匀分布，即

$$f_X(x) = \begin{cases} \frac{1}{\theta}, 0 < x \leqslant \theta \\ 0, \ x > \theta \end{cases}$$

试计算θ的最大似然估计。令n个观测$X_1, \cdots, X_n$的一个特定实现为$\boldsymbol{x} = [x_1, \cdots, x_n]^{\mathrm{T}}$，并令$x_m \triangleq \max(x_1, \cdots, x_n)$，似然函数为

$$L(\theta) = \begin{cases} \frac{1}{\theta^n}, x_m \leqslant \theta \\ 0, \quad 其他 \end{cases}$$

显然，最大化L必须使θ'的估计尽可能小，但θ'不能小于x_m，因此，θ'就等于x_m，最大似然估计为

$$\hat{\theta} = \max(X_1, \cdots, X_n) \tag{6.7-5}$$

当$n=2$时$\hat{\theta}$的累积密度函数为

$$F_{\hat{\theta}}(\alpha) = F_{X_1}(\alpha)F_{X_2}(\alpha) = F_X^2(\alpha) \tag{6.7-6}$$

任意n时$\hat{\theta}$的累积密度函数和概率密度函数计算留给读者练习。

例 6.7-5 考虑正态概率密度函数

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \quad -\infty < x < \infty$$

n个实现的对数似然函数为

$$\begin{aligned} \bar{L}(\mu,\sigma) \triangleq \log L = -\frac{n}{2}\log 2\pi - n\log\sigma \\ -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \end{aligned} \tag{6.7-7}$$

令

$$\frac{\partial \bar{L}}{\partial \mu} = 0 \quad \frac{\partial \bar{L}}{\partial \sigma} = 0$$

得到联立方程

$$\sum_{i=1}^{n}(x_i-\mu) = 0 \tag{6.7-8}$$

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \tag{6.7-9}$$

根据式(6.7-8)可得

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{6.7-10}$$

根据式(6.7-9)，并应用式(6.7-10)的结果可得

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 \tag{6.7-11}$$

最大似然估计有很多优良的性质包括均方误差一致性和不变性。不变性指的是如果 $\hat{\theta}$ 是 θ 的最大似然估计，则 $h(\hat{\theta})$ 是 $h(\theta)$ 的最大似然估计。然而，如例 6.7-5 所示[参见式(6.7-11)]最大似然估计不一定是无偏的。我们用一个不变性的例题结束本节。

例 6.7-6 考虑正态随机变量的 n 个观测，假定已知其均值为零。方差的最大似然估计为 $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2$，标准差 σ 是方差的均方根。则标准差的最大似然估计为方差最大似然估计的均方根，即 $\hat{\sigma} = \left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)^{1/2}$。

6.8 排序，百分点，参数与非参数统计

取累积分布函数为 $F_X(x)$ 的一类随机变量 X(X 有时称为集合)的 n 个独立同分布观测 $X_1, X_2, \cdots, X_n$，样本的联合概率密度函数为 $f_X(x_1)\times\cdots\times f_X(x_n)$，$-\infty < x_i < \infty$，$i=1, \cdots, n$，将 X_i，$i=1, \cdots, n$ 按大小(符号数值)排序后得到排序的样本 $Y_1, Y_2, \cdots, Y_n$，即 $-\infty < Y_1 < Y_2 < \cdots < Y_n < \infty$，例如序列 3，-2，-9，4 排序后为 -9，-2，3，4。若序列 $X_1, \cdots, X_{20}$ 是由 $X:N(0,1)$ 的 n 个观测生成，则 $Y_1 > 0$ 的可能性不大，因为这需要其他 19 个 Y_i，$i=2, \cdots, 20$ 都大于零，即所有样本都在正态曲线的右侧，这种可能性为 $(1/2)^{20}$。同样，$Y_{20} < 0$ 也是不太可能的，因为这需要其他 19 个 Y_i，$i=1, \cdots, 19$ 都小于零。如 5.3 节所示，排序样本 $Y_1, Y_2, \cdots, Y_n$ 的联合概率密度函数为 $n!\ f_X(y_1)\times\cdots\times f_X(y_n)$，$-\infty < y_1 < y_2 < \cdots < y_n < \infty$ 或零。我们区分排序和排名是排名通常给排序单元指派一个值，例如，大多数人会将骨折的疼痛排序在感冒引起的喉咙疼痛前面，但若一个内科医生要病人将疼痛按 0~10 排名，骨折的疼痛可能排名第 8 位或第 9 位，而喉咙疼痛可能排名第 3 位或第 4 位。

下面考虑百分点的概念，在前面很多地方讨论中已应用过此概念，我们将在此详细阐述。假定一大群人的 IQ(智商)分布为 $N(100, 100)$，即均值为 100，标准差为 10。显然，正态近似是在一定范围内有意义的，因为没有人的 IQ 会是 1000 或是 -10。IQ 测试本身也是在一定范围内有意义的，对于特别聪明或严重智力低下的也难以给出精确的数值，有时将这两类人叫做“脱离 IQ 范围”。现阶段 IQ 测试依然作为解决问题能力的指标广泛应用，假定 IQ 测试结果表明小孩在应试者中排名是 93 百分点，有资格进入所选择的学校，我们怎么在 n 个学生中定为 93 的百分点呢？

定义(百分点) 给定累积密度函数 $F_X(x)$ 的随机变量 X，X 的 u 百分点是满足 $F_X(x_u) = u$ 的 x_u 值。若函数 F_X 是处处连续且可微的，则 $x_u = F_X^{-1}(u)$，其中 F_X^{-1} 是 F_X 的逆函数，满足 $F_X^{-1}(F_X(x_u)) = x_u$。图 6.8-1 画出了一个累积概率密度及其逆函数。按惯例，我们不加区别地用 x_u 或 $100\times x_u$ 来表示 x_u 百分点。

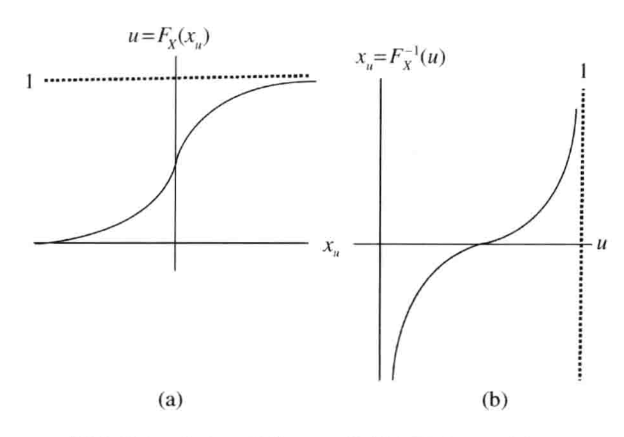


图 6.8-1 (a)u 对比 x_u；(b)逆函数 x_u 对比 u

观测 在累积分布函数为 $F_{SN}(z)$ 的 $X:N(0,1)$ 的特定情况，我们用符号 z_u(或 $100\times z_u$)表示 X 的 u 百分点。若 $X:N(\mu,\sigma^2)$，则 X 的 u 百分点 x_u 与 z_u 的关系满足

$$x_u=\mu+z_u\sigma \tag{6.8-1}$$

例 6.8-1 (x_u 与 z_u 的关系)若 $X:N(u,\sigma^2)$，证明 $x_u=\mu+z_u\sigma$。证明如下：
我们有

$$\begin{aligned}F_X(x_u)=u&=\left(2\pi\sigma^2\right)^{-1/2}\int_{-\infty}^{x_u}\exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)\mathrm{d}x\\&=(2\pi)^{-1/2}\int_{-\infty}^{(x_u-\mu)/\sigma}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}z\\&\triangleq(2\pi)^{-1/2}\int_{-\infty}^{z_u}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}z\end{aligned}$$

最后一行是 $F_{SN}(z_u)$，即标准正态随机变量的累积分布函数。因此 $x_u=\mu+z_u\sigma$。我们将此结论用于前面提到的 IQ 问题。根据数据有 $F_X(x_u)=0.93=F_{SN}(z_u)$，从正态累积分布函数表或从误差函数($\mathrm{erf}(z_u)=F_{SN}(z_u)-0.5$)表中可得 z_u，即 $z_u\approx1.48$，根据 $x_u=\mu+z_u\sigma=100+1.48(10)$，得到的 IQ 93 百分点为 115。

集合的中值与均值

集合 X 的中值是满足 $F_X(x_{0.5})=0.5$[①] 的点 $x_{0.5}$，而 X 的均值写成 μ_X，定义为 $\mu_X=\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x$。中值和均值并不需要一致，例如，在 $f_X(x)=\lambda\mathrm{e}^{-\lambda x}u(x)$ 的指数分布中，可得 $\mu_X=1/\lambda$ 及 $x_{0.5}=0.69/\lambda$。计算 X 的均值需要 $f_X(x)$，但这经常是不知道的。中值仅仅是一个点 $x_{0.5}$ 满足 $P[X\leqslant x_{0.5}]$，而均值看起来是一个相对抽象的参数。给定 X 的 n 个独立同分布观测 $X_1, X_2, \cdots, X_n$，用均值估计函数(MEF) $\hat{\mu}_X=n^{-1}\sum_{i=1}^{n}X_i$ 来估计 μ_X，在大数据量时

① 当事件 $\{X=x_{0.5}\}$ 概率为零，事件 $\{X<x_{0.5}\}$ 及 $\{X>x_{0.5}\}$ 概率都为 0.5，这就是经常所听到的描述：中值是“集合中一半在上，一半在下的点”。中值作为 50 百分点，它包含事件 $\{X=x_{0.5}\}$，因此，描述应该修改为：“中值是集合中一半在其点或在下”。中值是整个集合的一个特征参数，随机样本的中值仅仅是真实中值的估计。

是无偏且一致估计量。实际上由于 MEF $\hat{\mu}_X$ 的形式简单，并且在 σ_X^2 有限且大 n 时(参见大数定理)，有 $\hat{\mu}_X \to \mu_X$，才使得均值应用广泛。MEF 的实现直观上是有趣的，它给出了数据的着重描述。

例 6.8-2 (薪水的中值和均值)考虑一个国家中一半的工人一年的收入少于 \$10 000，而另一半多于 \$10 000，我们将 \$10 000 作为年收入的中值。假定在年收入少于 \$10 000 的人中，年收入的均值是 \$8000，在年收入多于 \$10 000 的人中，年收入的均值是 \$100 000，则整个国家的收入均值为 \$54 000。依你的判断，哪个数据能更好体现国家的经济？国家运行合理(不合理)你会使用哪一个指标？

例 6.8-3 (二项式的中值和均值的不同)我们对二项式情况做了一些细致的观察，发现均值和中值不一致。例如，当 $n=5$，均值是 2.5，但中值是 2。然而，当 n 很大时，中值和均值接近，且中值可用均值来估计。不加证明的结论是：均值和中值的差别与 $(p(1-p))^n$ 成比例，当 $n \to \infty$ 为任意小。

参数与非参数统计

已知或假定密度、分布或概率分布函数，然后根据这些信息计算概率、估计参数、做出判决称为参数统计。典型的参数统计中，我们会假定集合密度的形式，比如正态分布，然后去估计一些分布相关的未知参数，比如均值 μ_X。在第 7 章中，将参数统计应用于假设检验。参数统计很大程度上是基于中心极限定理，表明大量的独立同分布观测的分布趋近于正态累积密度函数。

集合的特性和参数的估计不依赖于集合分布的知识或形式的假定称为分布无关或非参数统计。仅仅基于观测而没有分布假定的统计有时称为鲁棒的，这是因为根据观测得到的定理或结论不会因为分布形式的不同而改变。均值和标准差在描述集合的中心和散布时是十分有用的参量，而在非参量情形中，中值和范围扮演相似的角色。为从 $X_1, X_2, \cdots, X_n$ 中估计中值，我们将它们按大小排列为 $Y_1 < Y_2 < \cdots Y_n$，用样本中值估计器估计 $x_{0.5}$ 为

$$\hat{Y}_{0.5} = \begin{cases} Y_{k+1} & \text{若 } n \text{ 为奇数，则 } n=2k+1 \\ 0.5(Y_k + Y_{k+1}) & \text{若 } n \text{ 为偶数，则 } n=2k \end{cases} \tag{6.8-2}$$

样本中值不是 $x_{0.5}$ 的无偏估计，但在 n 很大时近似无偏。非参量情形的散布用 50% 范围描述，即 $\Delta x_{0.50} \triangleq x_{0.75} - x_{0.25}$，或用 90% 范围描述，即 $\Delta x_{0.90} \triangleq x_{0.95} - x_{0.05}$，或用其他合适的范围描述，这都将根据观测值进行估计。

例 6.8-4 (百分点的内插)符号 $\alpha \sim \beta$ 表示 α 估计 β，有 $Y_3 \sim x_{0.273}$，$Y_4 \sim x_{0.364}$，应用内插公式

$$Y_4 + \frac{(Y_4 - Y_3)(0.3 - 4/11)}{1/11} \sim x_{0.3}$$

图 6.8-2 显示了样本间的内插。

百分点的置信区间

下面我们讨论一个与百分点排序统计相关的基本结论。模型依然是累积概率密度为 $F_X(x)$ 的随机变量 X 的 n 个独立同分布观测 $X_1, X_2, \cdots, X_n$，并应用标记 $P[X_i \leqslant x_u] \triangleq u$，下面将样本按符号数值大小排序为 $Y_1 < Y_2 < \cdots < Y_n$。在此提醒读者：若一个实现的集合 X_i，$i=$

1，…,5 为 $x_1=7$，$x_2=-2$，$x_3=7.2$，$x_4=1$，$x_5=3$，则排列后的实现 Y_i，$i=1$，…，5 为 $y_1=-2$，$y_2=-1$，$y_3=3$，$y_4=7$，$y_5=7.2$。从$\{Y_i\}$和$\{X_i\}$的下标看，可以得到明显的结论，即事件$\{Y_k<x_u\}$表明在$\{X_i\}$中至少有 k 个值小于 x_u，可能有更多，但不会少。因为$\{X_i\}$是独立同分布的，可以用二项式概率公式计算 $P[Y_k<x_u]$为

$$P[Y_k < x_u] = P[\{X_i\}\text{中至少有}\,k\,\text{个值小于}\,x_u]$$
$$= \sum_{i=k}^{n}\binom{n}{i}u^i(1-u)^{n-i} \tag{6.8-3}$$

下面考虑事件$\{Y_{k+r}>x_u\}$，由于 Y_{k+r}是$\{X_i\}$排序中第$(k+r)$个单元，则$\{X_i\}$中至少有 $n-(k+r)+1$ 个值大于 x_u，等效地，$\{X_i\}$中有不超过 $k+r-1$ 个值小于 x_u。即

$$P[Y_{k+r} > x_u] = P[\{X_i\}\text{中有不超过}\,k+r-1\,\text{个值小于}\,x_u]$$
$$= \sum_{i=0}^{k+r-1}\binom{n}{i}u^i(1-u)^{n-i} \tag{6.8-4}$$

事件$\{Y_{k+r}>x_u\}$和$\{Y_k<x_u\}$的并集为事件$\{Y_k<x_u<Y_{k+r}\}$，其概率

$$P[Y_k < x_u < Y_{k+r}] = \sum_{i=k}^{k+r-1}\binom{n}{i}u^i(1-u)^{n-i} \tag{6.8-5}$$

与$f_X(x)$是相互独立的。式(6.8-5)给出的结论是非参量统计中的重要结论并将在下面得到应用。

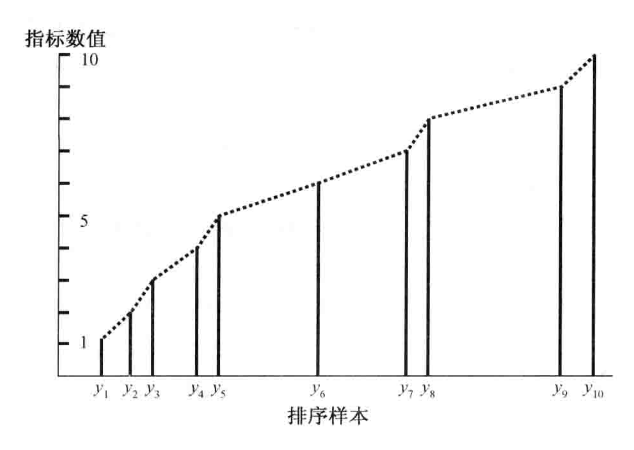


图 6.8-2　从 10 个排序的样本估计百分点范围显示为样本间线性内插。为获取估计的百分点，采用指标数值并乘以100/11，这样，对于初始近似，90百分点从y_{10}估计，而9百分点从y_1估计，y_8-y_2覆盖了50百分点范围

例 6.8-5　（中值的 95% 置信度需要的样本数）我们寻求随机区间$[Y_1, Y_n]$的端点 Y_1，Y_n，满足事件$\{Y_1<x_{0.5}<Y_n\}$发生的概率是 0.95。此处 $Y_1\triangleq\min(X_1, X_2, \cdots, X_n)$，$Y_n\triangleq\max(X_1, X_2, \cdots, X_n)$，则实际上 n 应该取多大呢？

答案是通过计算

$$P[Y_1 < x_{0.5} < Y_n] = \sum_{i=1}^{n-1}\binom{n}{i}(1/2)^n \approx 0.95$$

可得当 $n=5$，$P[Y_1<x_{0.5}<Y_5]\approx0.94$，随机区间$[Y_1, Y_n]$覆盖50% 的百分点的概率随 n 的变化规律如图 6.8-3 所示。

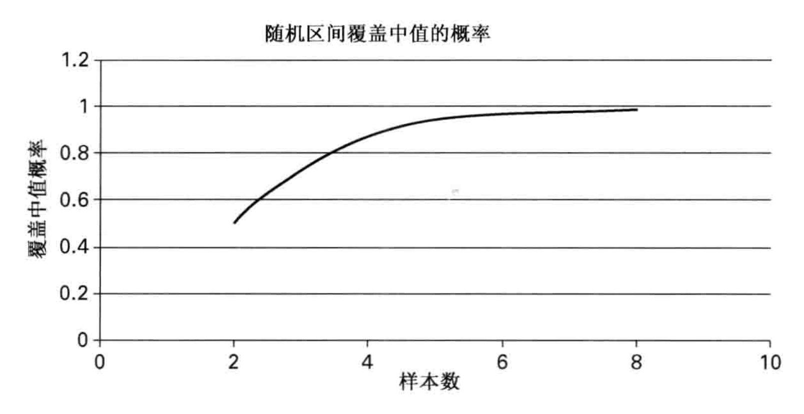


图6.8-3 事件 $\{Y_1 < x_{0.5} < Y_n\}$ 覆盖中值的概率与 n 值关系

例6.8-6 （$x_{0.33}$位于哪一组排序样本间?）有一组排序的样本 $\{Y_1, Y_2, \cdots, Y_n\}$，希望找到一个样本对 $\{Y_i, Y_{i+1}, i=1, \cdots, n-1\}$ 覆盖33.33百分点的概率最大，33.33百分点由 $u=1//3=F_X(x_{0.33})$ 给出。假定 $n=10$，根据式(6.8-5)可得

$$P\left[Y_k < x_{0.33} < Y_{k+1}\right] = \frac{10!}{k!(10-k)!}(1/3)^k(2/3)^{10-k}, k=1,\cdots,9$$

结果绘在图6.8-4中。显然间隔 $[Y_3, Y_4]$ 覆盖 $x_{0.33}$ 的可能性最大，事件 $\{Y_3 < x_{0.33} < Y_4\}$ 的概率为0.26。

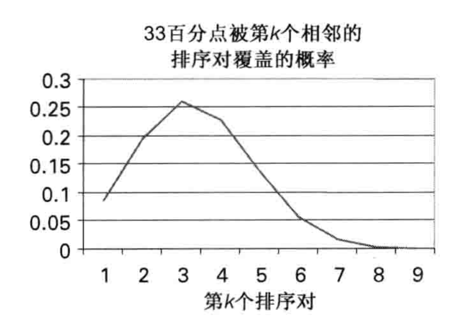


图6.8-4 在所有成对区间 $[Y_k, Y_{k+1}]$ 中，区间 $[Y_3, Y_4]$ 覆盖 $x_{0.33}$ 可能性最大

当 n 很大时中值的置信区间

若 n 足够大，以至于二项式分布可以用正态分布近似，我们有

$$P[\alpha \leqslant S_n \leqslant \beta] \approx \frac{1}{\sqrt{2\pi}} \int_{\alpha_n}^{\beta_n} \exp\left[-\frac{1}{2}y^2\right] \mathrm{d}y \tag{6.8-6a}$$

其中

$$P[\alpha \leqslant S_n \leqslant \beta] = \sum_{i=\alpha}^{\beta} \binom{n}{i} p^i (1-p)^{n-i}$$

$$\alpha_n \triangleq \frac{\alpha - np - 0.5}{\sqrt{np(1-p)}} \tag{6.8-6b}$$

$$\beta_n \triangleq \frac{\beta - np + 0.5}{\sqrt{np(1-p)}}$$

应用此结果于就近的问题，有

$$P[Y_r < x_{0.5} < Y_{n-r+1}] = \sum_{i=r}^{n-r} \binom{n}{i} (1/2)^n \tag{6.8-7}$$

在应用中中值定义为 $u = F_X(x_{0.5}) = 1/2$。下标的选择确保置信区间开始从底部的第 r 个位置计算，也就是说，由首个位置开始，结束于从顶部开始往回计算 r 个观测位置。举例来说，如果 $n=10$ 时 95% 置信区间得到 $r=3$，置信区间从第 3 个观测开始，结束于第 8 个观测，两个点均是从底部或顶部开始的第 3 个位置，即 1，2，3(Y_3)和 10，9，8(Y_8)，结果可表示为 $P[Y_3 < x_{0.5} < Y_8] = 0.95$。

从二项式求和的式(6.8-7)注意到，其均值为 $n/2$，标准差为$\sqrt{n}/2$，而二项式求和式(6.8-7)的 95% 置信区间正态近似为

$$\sum_{i=r}^{n-r} \binom{n}{i} (1/2)^n \approx \frac{1}{\sqrt{2\pi}} \int_{\alpha_n}^{\beta_n} \exp[-\frac{1}{2}x^2]\mathrm{d}x = 0.95$$

由标准正态分布(或误差函数)表可得 $\alpha_n = -1.96$，$\beta_n = 1.96$。根据式(6.8-6b)有

$$1.96 = \frac{n - r - n/2 + 0.5}{\sqrt{n}/2}$$

$$-1.96 = \frac{r - n/2 - 0.5}{\sqrt{n}/2}$$

可得 $r = (n/2) - 1.96\sqrt{n}/2 + 0.5$。如果 r 不是整数，则用$[r]$代替 r，$[r]$为小于等于 r 的最大整数。

例 6.8-7 ($n=20$ 时中值的 95% 置信区间)随机变量 X 的 20 个观测为$\{X_i, i=1, \cdots, 20\}$，根据其符号数值排序为 $Y_1 < Y_2 < \cdots < Y_n$，由 $r = (n/2) - 1.96\sqrt{n}/2 + 0.5$ 可得 $r=6.12$ 和$\lfloor r \rfloor = 6$，则有 $P[Y_6 < x_{0.5} < Y_{15}] \geqslant 0.95$。

6.9 向量均值和协方差阵的估计①

令 $\boldsymbol{X}_1 \triangleq [X_i, \cdots, X_p]^{\mathrm{T}}$ 为 p 单元的随机向量，概率密度函数为$f_X(x)$，令 $X_i, \cdots, X_n$ 为 X 的 n 个观测，即 X_i，$i=1, \cdots, n$ 是由$f_X(x)$给出的，则 X_i，$i=1, \cdots, n$ 为独立同分布的随机向量，概率密度函数为$f_X(x_i)$。下面我们给出如何估计

$$\text{(i) } \boldsymbol{\mu_X} \triangleq E[\boldsymbol{X}] = [\mu_1, \cdots, \mu_p]^{\mathrm{T}}$$

其中

$$\mu_j \triangleq E[X_j] \qquad j = 1, \cdots, p$$

及

$$\text{(ii) } \boldsymbol{K_{XX}} \triangleq E[(\boldsymbol{X} - \boldsymbol{\mu_X})(\boldsymbol{X} - \boldsymbol{\mu_X})^{\mathrm{T}}]$$

向量和矩阵$\boldsymbol{\mu_X}$ 和 $\boldsymbol{K_{XX}}$在信号处理应用中是非常有用的，它们也是多维正态分布的首要特征[6-5]。协方差阵 $\boldsymbol{K_{XX}}$通常是满秩、正定、实对称的矩阵，矩阵的这些性质是众所周知的[6-6]，并在估计中将得到应用。

① 这一节和下一节对首次阅读的读者可跳过。

μ 的估计

考虑给定的 p 维向量估计 $\hat{\boldsymbol{\Theta}}$

$$\hat{\boldsymbol{\Theta}} \triangleq \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i \tag{6.9-1}$$

可知 $\hat{\boldsymbol{\Theta}}$ 是 $\boldsymbol{\mu}$ 的无偏和一致估计。我们将观测整理在表6.9-1中。

表 6.9-1　观测数据

	$\boldsymbol{X}_1 \ldots \boldsymbol{X}_i \ldots \boldsymbol{X}_n$
Y_1	X_{11}　X_{i1}　X_{n1}
⋮	⋮
Y_j	X_{ij}　p 行
⋮	⋮
Y_p	X_{1p}　X_{np}
	n 列

$\boldsymbol{Y}_j$ 是估计向量 $\boldsymbol{\mu}$ 第 j 个 μ_j 的充分必要条件

在表6.9-1中，X_{ij}是随机向量 $\boldsymbol{X}_i$ 的第j个单元，Y_j，$j=1,\cdots,p$ 是随机向量 $\boldsymbol{X}$ 的第j个单元的 n 个独立同分布的观测。从标量情形已知

$$\hat{\Theta}_j \triangleq \frac{1}{n}\sum_{i=1}^{n}X_{ij} \triangleq \hat{\mu}_j \quad j=1,\cdots,p \tag{6.9-2}$$

对于 $\mu_j \triangleq E[X_{ij}]\,i=1,\cdots,n$ 而言是无偏且一致的，进而有 $\hat{\boldsymbol{\Theta}} \triangleq [\hat{\Theta}_1,\cdots,\hat{\Theta}_P]^{\mathrm{T}}$ 是 $\boldsymbol{\mu}$ 的无偏且一致的。向量 $\boldsymbol{Y}_j$ 包括估计 μ_j 的所有信息，因此 $E[\boldsymbol{Y}_j]=\mu_j\boldsymbol{i}$，其中 $\boldsymbol{i}\triangleq[1,1,\cdots,1,1]^{\mathrm{T}}$。

当 $\boldsymbol{X}$ 为正态时，$\hat{\boldsymbol{\Theta}}$ 也是正态的。即使 $\boldsymbol{X}$ 不是正态的，有中心极限定理，$\hat{\boldsymbol{\Theta}}$ 在 n 很大时也是趋近正态的(参见定理4.7-1)。

协方差 *K* 的估计

如果均值 $\boldsymbol{\mu}$ 已知，则估计

$$\hat{\boldsymbol{\Theta}} \triangleq \frac{1}{n}\sum_{i=1}^{n}[\boldsymbol{X}_i-\boldsymbol{\mu}][\boldsymbol{X}_i-\boldsymbol{\mu}]^{\mathrm{T}} \tag{6.9-3}$$

是 $\boldsymbol{K}$ 的无偏估计。而均值通常由样本均值 $\hat{\boldsymbol{\mu}}$ 估计给出，则估计

$$\hat{\boldsymbol{\Theta}} \triangleq \frac{1}{n-1}\sum_{i=1}^{n}[\boldsymbol{X}_i-\hat{\boldsymbol{\mu}}][\boldsymbol{X}_i-\hat{\boldsymbol{\mu}}]^{\mathrm{T}} \tag{6.9-4}$$

是 $\boldsymbol{K}_{XX}$的无偏估计。为证明此结论需要下一点工夫。首先观察 $\hat{\boldsymbol{\Theta}}$ 的对角线上单元形式

$$S_{jj} \triangleq \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij}-\hat{\mu}_j)^2 \tag{6.9-5}$$

从单变量情形我们已知 $\sigma_j^2 \triangleq E[(X_j-\mu_j)^2]$是无偏的。其次考虑序列($l\neq m$)

$$X_{1l}+X_{1m},X_{2l}+X_{2m},\cdots,X_{nl}+X_{nm} \tag{6.9-6}$$

为 n 个独立同分布观测 $Z_{lm}^{(i)}$，其中单变量随机变量 $Z_{lm}\triangleq X_l+X_m$ 均值为 $\mu_l+\mu_m$，方差为

$$\begin{aligned}\operatorname{Var}[Z_{lm}] &= E[(X_l-\mu_l)+(X_m-\mu_m)]^2 \\ &= \sigma_l^2+\sigma_m^2+2K_{lm}\end{aligned} \tag{6.9-7}$$

其中 $K_{lm}\triangleq E[(X_l-\mu_l)(X_m-\mu_m)]$是 $\boldsymbol{K}_{XX}$的第 lm 个单元。最后，考虑

$$\hat{\boldsymbol{\Theta}}_{lm} \triangleq \frac{1}{n-1}\sum_{l=1}^{n}[Z_{lm}^{i}-(\hat{\mu}_l+\hat{\mu}_m)]^2 \tag{6.9-8}$$

根据式(6.8-15)，上式是 $\sigma_l^2+\sigma_m^2+2K_{lm}$的无偏估计。将式(6.9-8)加以推广并应用 $Z_{lm}\triangleq X_l+$

X_m，可得

$$\hat{\boldsymbol{\Theta}}_{lm} \triangleq \frac{1}{n-1}\sum_{i=1}^{n}[(X_{il}-\hat{\mu}_l)+(X_{im}-\hat{\mu}_m)]^2$$
$$= \frac{1}{n-1}\sum_{i=1}^{n}(X_{il}-\hat{\mu}_l)^2 + \frac{1}{n-1}\sum_{i=1}^{n}(X_{im}-\hat{\mu}_m)^2 \tag{6.9-9}$$
$$+ \frac{2}{n-1}\sum_{i=1}^{n}(X_{il}-\hat{\mu}_l)(X_{im}-\hat{\mu}_m)$$

在式(6.9-9)中，第一项是σ_l^2的无偏估计，第二项是σ_m^2的无偏估计，三项之和是式(6.9-8)的$\sigma_l^2+\sigma_m^2+2K_{lm}$的无偏估计，因此可得出结论

$$S_{lm} \triangleq \frac{1}{n-1}\sum_{i=1}^{n}(X_{il}-\hat{\mu}_l)(X_{im}-\hat{\mu}_m) \tag{6.9-10}$$

是$K_{lm}(=K_{ml})$的无偏估计。既然式(6.9-4)中$\hat{\boldsymbol{\Theta}}$的每一项都是$\boldsymbol{K}_{XX}$对应项的无偏估计，因此$\hat{\boldsymbol{\Theta}} \triangleq \boldsymbol{K}_{XX}$是$\boldsymbol{K}_{XX}$无偏估计。

通过分析单变量情形并假定四阶矩存在，可以得出式(6.9-4)对$\boldsymbol{K}_{XX}$估计的每一项的一致性，因此无须指定分布，式(6.9-1)和式(6.9-4)分别是$\boldsymbol{\mu}_X$和$\boldsymbol{K}_{XX}$的无偏且一致估计量。

当$\boldsymbol{X}$为正态，$\hat{\boldsymbol{K}}_{XX}$结构上为复概率形式，称为 Wishart 分布(参见参考文献[6-6]的 126 页)。更一般地，当$\boldsymbol{X}$的概率密度函数形式已知，就可以采用最大似然方法估计$\boldsymbol{\sigma}_X^2$，$\boldsymbol{\mu}_X$和$\boldsymbol{K}_{XX}$等参数。尽管不都是，最大似然方法还是具有一些优良的性质。下个例子将说明均值的最大似然估计不是最小均方估计。

例 6.9-1 (参见参考文献[6-5]的 21 页)考虑式(6.8-3)的样本均值估计器，即

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}X_i$$

可知估计器是均值μ'的最大似然估计。问：标量估计器$\hat{\boldsymbol{\Theta}} \triangleq a\hat{\mu}$中常量$a$取什么值能得到$\mu$的最小均方估计？其中$X_i$，$i=1$，…，$n$是独立同分布的随机变量且有$E[X_i]=\mu$和$\mathrm{Var}[X_i]=\sigma^2$。

解 通过下式求解最小a值

$$E[a\hat{\mu}-\mu]^2 \tag{6.9-11}$$

显然$\hat{\mu}$是μ的无偏估计，很难相信当$a\neq 1$的$\hat{\boldsymbol{\Theta}}$，即得到一个有偏估计，会有比$\hat{\boldsymbol{\Theta}}=\mu$时更小的均方误差。

对于任意估计量$\hat{\boldsymbol{\Theta}}$，估计$\mu$的均方误差为

$$\begin{aligned} E[(\hat{\boldsymbol{\Theta}}-\mu)^2] &= E[\{(\hat{\boldsymbol{\Theta}}-E[\hat{\boldsymbol{\Theta}}])+(E[\hat{\boldsymbol{\Theta}}]-\mu)\}^2] \\ &= \mathrm{Var}[\hat{\boldsymbol{\Theta}}]+(E[\hat{\boldsymbol{\Theta}}]-\mu)^2 \end{aligned} \tag{6.9-12}$$

如果$\hat{\boldsymbol{\Theta}}$是无偏的，则最后一项即偏差的平方项(参见定义 6.8-2)为零。对于此例来说，$\hat{\boldsymbol{\Theta}}=a\hat{\mu}$，因此

$$\begin{aligned} E[(\hat{\boldsymbol{\Theta}}-\mu)^2] &= a^2\mathrm{Var}[\hat{\mu}]+(a\mu-\mu)^2 \\ &= \frac{a^2\sigma^2}{n}+(a-1)^2\mu^2 \end{aligned} \tag{6.9-13}$$

为得到最小均方误差估计，将式(6.9-13)对 a 微分并令等于零，得到最佳值 $a=a_0$，即

$$a_0=\frac{\mu^2}{(\sigma^2/n)+\mu^2}=\frac{n}{(\sigma^2/n)+n} \tag{6.9-14}$$

6.10 向量参数的线性估计①

现实中许多测量问题可用如下模型描述：

$$y(t)=\int_T h(t,\tau)\theta(\tau)\mathrm{d}\tau+n(t) \tag{6.10-1}$$

其中 $y(t)$ 是观测或测量值，T 是积分设定，$\theta(\tau)$ 是未知参数函数，$h(t,\tau)$ 是联系参数函数到测量系统的特性函数，与 $\theta(\tau)$ 独立，$n(t)$ 是不可避免的测量噪声。为计算方便，式(6.10-1)需要转换到离散形式

$$\boldsymbol{Y}=\boldsymbol{H\theta}+\boldsymbol{N} \tag{6.10-2}$$

其中 $\boldsymbol{Y}$ 是由 Y_i，$i=1,\cdots,n$ 组成的 $n\times1$ 的观测向量，$\boldsymbol{H}$ 是已知的 $n\times k$ 矩阵($n>k$)，$\boldsymbol{\theta}$ 是未知的 $k\times1$ 参数向量，N 是由第 i 个观测 Y_i 对应的噪声误差 N_i，$i=1,\cdots,n$ 所组成的 $n\times1$ 的随机向量，不失一般性，假定 $E[\boldsymbol{N}]=\boldsymbol{0}$②。

式(6.10-2)是线性模型，回答以下问题：如果约束估计 $\hat{\boldsymbol{\Theta}}$ 是 $\boldsymbol{Y}$ 的线性函数，我们如何从观测 $\boldsymbol{Y}$ 中得到 $\boldsymbol{\theta}$ 的一个好估计。线性函数意味着

$$\hat{\boldsymbol{\Theta}}=\boldsymbol{BY} \tag{6.10-3}$$

其中 $\boldsymbol{B}$ 与 $\boldsymbol{Y}$ 独立，需要计算得到。这个问题很具有实际意义，它是参数估计理论中最基本的问题，在许多教材中有详细描述，如 Kendall 和 Stuart[6-8]，Lewis 和 Odell[6-9]。它也是随机向量概率理论的直接应用，对于后续章节各专题的理解非常有帮助。

在用式(6.10-3)计算矩阵 $\boldsymbol{B}$ 之前，首先介绍矩阵微积分中的一些结论。

标量函数对向量的微分　令 $q(\boldsymbol{x})$ 为向量 $\boldsymbol{x}=[x_1,\cdots,x_n]^{\mathrm{T}}$ 的标量函数，则有

$$\frac{\mathrm{d}q(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}\triangleq\left[\frac{\partial q}{\partial x_1},\cdots,\frac{\partial q}{\partial x_n}\right]^{\mathrm{T}} \tag{6.10-4}$$

因此，$q(\boldsymbol{x})$ 对 $\boldsymbol{x}$ 的微分是一列向量，其第 i 单元为 $q(\boldsymbol{x})$ 对 x_i 的偏导。

二次型的微分　令 $\boldsymbol{A}$ 为一实对称 $n\times n$ 矩阵，$\boldsymbol{x}$ 为一任意 n 维向量，则二次型

$$q(\boldsymbol{x})\triangleq\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Ax}$$

对 $\boldsymbol{x}$ 的微分为

$$\frac{\mathrm{d}q(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}}=2\boldsymbol{Ax} \tag{6.10-5}$$

式(6.10-5)证明如下：

$$\begin{aligned}q(\boldsymbol{x})&=\sum_{i=1}^{n}\sum_{j=1}^{n}x_ia_{ij}x_j\\&=\sum_{i=1}^{n}x_i^2a_{ii}+\sum_{i\neq j}^{n}\sum^{n}a_{ij}x_ix_j\end{aligned}$$

---

① 这一节对首次阅读的读者可跳过。

② 符号 0 此处表示零向量，也就是说，向量的各单元都是零。

又

$$\frac{\partial q(\boldsymbol{x})}{\partial x_k} = 2x_k a_{kk} + 2\sum_{i\neq k} a_{ki}x_i$$

$$= 2\sum_{i=1}^{n} a_{ki}x_i$$

即

$$\frac{\mathrm{d}q(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x} \tag{6.10-6}$$

标量积的微分 令 $\boldsymbol{a}$ 和 $\boldsymbol{x}$ 为两个 n 维向量，由 $y=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$，有

$$\frac{\mathrm{d}y}{\mathrm{d}\boldsymbol{x}} = \boldsymbol{a} \tag{6.10-7}$$

令 $\boldsymbol{x}$, $\boldsymbol{y}$ 为两个 $\boldsymbol{n}$ 维向量，$\boldsymbol{A}$ 为 $n\times n$ 矩阵，则由 $q\triangleq\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$，有

$$\frac{\partial q}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \tag{6.10-8}$$

回到式(6.10-2)

$$\boldsymbol{Y} = \boldsymbol{H}\theta + \boldsymbol{N}$$

并假定(回想 $E[\boldsymbol{N}]=\boldsymbol{0}$)

$$\boldsymbol{K} \triangleq E[\boldsymbol{N}\boldsymbol{N}^{\mathrm{T}}] = \sigma^2\boldsymbol{I} \tag{6.10-9}$$

其中 $\boldsymbol{I}$ 是单位矩阵。式(6.10-9)表明测量误差 N_i，$i=1$，…，n 是不相关的，其方差相同且等于 σ^2，这有时称为白噪声。

估计 $\boldsymbol{\theta}$ 的一个合理选择是寻找使下述定义的均方和 S 最小的 $\hat{\boldsymbol{\Theta}}$

$$S \triangleq (\boldsymbol{Y}-\boldsymbol{H}\hat{\boldsymbol{\Theta}})^{\mathrm{T}}(\boldsymbol{Y}-\boldsymbol{H}\hat{\boldsymbol{\Theta}}) \triangleq ||\boldsymbol{Y}-\boldsymbol{H}\hat{\boldsymbol{\Theta}}||^2 \tag{6.10-10}$$

注意到通过 $\|\boldsymbol{Y}-\boldsymbol{H}\hat{\boldsymbol{\Theta}}\|^2$ 最小化得到的与测量 $\boldsymbol{Y}$ 最拟合的 $\hat{\boldsymbol{\Theta}}$，就是通常所说的与数据的最小二乘拟合。因此通过使式(6.10-10)的 S 最小化得到的 $\hat{\boldsymbol{\Theta}}$ 称为最小二乘(LS)法，它是最小均方误差估计的一种形式。为计算关于 $\hat{\boldsymbol{\Theta}}$ 的最小 S，有

$$S = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} + \hat{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}\hat{\boldsymbol{\Theta}} - \hat{\boldsymbol{\Theta}}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{Y} - \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{H}\hat{\boldsymbol{\Theta}}$$

计算得到[应用式(6.10-4)、式(6.10-5)和式(6.10-8)]

$$\frac{\partial S}{\partial \hat{\boldsymbol{\Theta}}} = 0 = 2[\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}]\hat{\boldsymbol{\Theta}} - 2\boldsymbol{H}^{\mathrm{T}}\boldsymbol{Y}$$

可得(假定 $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}$ 是可逆的)

$$\hat{\boldsymbol{\Theta}}_{LS} = (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{Y} \tag{6.10-11}$$

将结果与式(6.10-3)比较可见，式(6.10-3)中的 $\boldsymbol{B}$ 如给定为 $\boldsymbol{B}_0\triangleq(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}$，即为最小二乘估计。式(6.10-11)为 $\boldsymbol{\theta}$ 的基于测量 $\boldsymbol{Y}$ 的最小二乘估计。

精明的读者注意到我们没有应用 $\boldsymbol{K}=\sigma^2\boldsymbol{I}$，事实上，在得到式(6.10-11)过程时，将 $\boldsymbol{Y}$ 看成确定性的，而通过将方程 $\boldsymbol{Y}=\boldsymbol{H}\boldsymbol{\theta}$ 中的系统简单求逆得到 $\hat{\boldsymbol{\Theta}}_{\mathrm{LS}}$[参见 *Lewis* 和 *Odell*(6-9, p.6)]，实际上，由式(6.10-11)给出的估计 $\hat{\boldsymbol{\Theta}}_{\mathrm{LS}}$并不一定是最佳的。而当噪声 $\boldsymbol{N}$ 的方差如式(6.10-9)所示时，$\hat{\boldsymbol{\Theta}}_{\mathrm{LS}}$事实上具有最佳性质。我们留给读者证明 $\hat{\boldsymbol{\Theta}}_{\mathrm{LS}}$是无偏且最小方差估计。

例6.10-1 给定下列数据：

$$6.2 = 3\theta + n_1$$

$$7.8 = 4\theta + n_2$$

$$2.2 = \theta + n_3$$

求 θ 的最小二乘估计。

解 数据可以写成如下形式：

$$\boldsymbol{y} = \boldsymbol{H}\theta + \boldsymbol{n}$$

其中 $\boldsymbol{y} = [6.2, 7.8, 2.2]^{\mathrm{T}}$ 是 $\boldsymbol{Y}$ 的实现，$\boldsymbol{H}$ 是 $[3, 4, 1]^{\mathrm{T}}$ 的列向量，$\boldsymbol{n} = [n_1, n_2, n_3]^{\mathrm{T}}$ 是 $\boldsymbol{N}$ 的实现。由 $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H} = \sum H_i^2 = 26$ 及 $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{y} = \sum_{i=1}^{3} H_i y_i = 52$，有

$$\hat{\theta}_{\mathrm{LS}} = (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{y} = \frac{\sum_{i=1}^{3} H_i y_i}{\sum_{i=1}^{3} H_i^2} = \frac{52}{26} = 2$$

例6.10-2 ([6-8, p.77.]) 令 $\boldsymbol{\theta} = [\theta_1, \theta_2]^{\mathrm{T}}$ 为待估计的二元参数向量，令 $\boldsymbol{H}$ 为 $n \times 2$ 系数矩阵，可以分解成列向量形式 $\boldsymbol{H} = (\boldsymbol{H}_1\boldsymbol{H}_2)$，其中 $\boldsymbol{H}_i$，$i = 1, 2$ 为 n 维向量。观测数据为 n 维向量 $\boldsymbol{Y}$，线性模型形式假定为

$$\boldsymbol{Y} = (\boldsymbol{H}_1\boldsymbol{H}_2)\boldsymbol{\theta} + \boldsymbol{N}$$

$\boldsymbol{\theta}$ 的最小二乘估计为

$$\hat{\boldsymbol{\Theta}}_{\mathrm{LS}} = \begin{bmatrix} \boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{H}_1 & \boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{H}_2 \\ \boldsymbol{H}_2^{\mathrm{T}}\boldsymbol{H}_1 & \boldsymbol{H}_2^{\mathrm{T}}\boldsymbol{H}_2 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{Y} \\ \boldsymbol{H}_2^{\mathrm{T}}\boldsymbol{Y} \end{bmatrix}$$

小结

在统计学的参数估计部分，我们对观测数据应用概率工具来估计联系概率函数的参数。本章开始强调了对感兴趣随机变量的独立同分布观测重要性，然后描述了如何根据观测来估计均值、方差等参数，特别是在正态分布情况，用置信区间的概念来对观测数目有限时如何对参数下一个有力(确信无疑)的结论，这样，我们可以说基于观测，真实的均值或方差或两者会接近100%置信度位于计算出的间隔内，我们研究了标准均值估计函数，其是无偏且一致的。

我们发现描述 T 随机变量概率行为的 t 分布，对分析正态随机变量在方差未知时的均值置信区间非常重要。

在估计正态随机变量方差时，卡方分布在分析方差的接近100%置信区间时非常有用，我们简要讨论了从排序观测中估计正态随机变量标准差的方法。

我们证明了置信区间能用于分析正态分布以外的参数，通过指数和伯努利分布举例证明。

讨论了基于哪一个参量最有可能产生观测数据来估计参数的方法，这个方法称为最大似然估计(MLE)，这是个十分有效的方法，但并不总是得到无偏或最小均方误差估计。

在本章后一部分，介绍了非参数方法的参数估计，也称为分布无关估计，它不假定特定分布来生成观测数据，在某种意义上可以说是鲁棒的。我们发现采用排序数据和二项式分布，可以获得非参数情形的一系列重要结论。

最后，将前面讨论的参数估计推广到向量情况。特别是，我们表明了从观测数据中估计向量均值和协方差阵中的各单元，对测量数据进行线性运算来估计向量参数在本章末尾进行了简要讨论。

习题

(带 * 号的习题带有一定的难度，需要花更多精力对教材进一步的学习。)

6.1 正态随机变量 X 的有 n(偶数)个独立同分布观测，标记为 $\{X_i,\ i=1,\ \cdots,\ n\}$。我们将序列重组为：$Y_{2i}=X_{2i-1}$，$Y_{2i-1}=X_{2i}$，$i=1,\ \cdots,\ n/2$，$\{Y_i\}$ 中各随机变量是否独立？

6.2 有 $X:N(0,1)$ 的三个独立同分布观测 X_i，$i=1,2,3$，计算 $f_{X_1X_2X_3}(x_1,\ x_2,\ x_3)$ 并与 $f_{X_1+X_2+X_3}(y)$ 相比较。

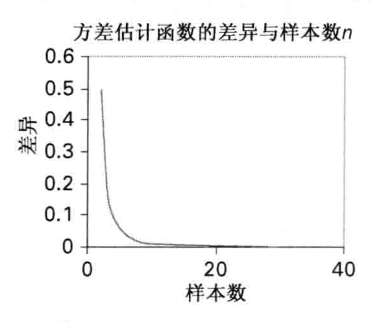
6.3 在一个发展中国家的村落，361 名村民受到埃博拉病毒威胁，其中 189 名村民死于病毒感染，计算受到埃博拉病毒威胁后死亡概率的 95% 置信区间，误差率是多少？

6.4 证明例 6.1-6 出现的多项式 $(p-\hat{p})^2-(9/n)p(1-p)=0$ 的根实际就是式(6.1-1)。

6.5 参见例 6.1-6，计算当 $\hat{p}$ 从 0 变化到 1 时的 $|p_1-p_2|$，在不同的 n 取不同值时计算，取 $n=0,\ 20,\ 30,\ 50$。

6.6 试描述如何测试得到投掷硬币时“正面”朝上概率的 95% 置信区间。

6.7 考虑式(6.3-3)和式(6.3-4)的方差估计函数，证明当 $n\geqslant 20$ 时，它们的差别变得极小。再现如下的曲线。



6.8 当 $X:N(1,\ 1)$ 时，计算 n 的函数 $P[\ |\hat{\mu}_X(n)-\mu_X|\leqslant 0.1]$。

6.9 绘出单位方差的正态随机变量均值 95% 置信区间的宽度与样本数 n 的图形。

6.10 证明伽马概率密度函数

$$f(x;\alpha,\beta)=\left(\alpha!\beta^{(\alpha+1)}\right)^{-1}x^{\alpha}\exp(-x/\beta),x>0;\ \alpha>-1,\beta>0$$

的 MGF 为 $M(t)=(1-\beta t)^{-(\alpha+1)}$。

6.11 $X:N(\mu,\sigma^2)$ 的 n 个独立同分布观测为 X_i，$i=1,\ \cdots,\ n$，令 $Y_i=\dfrac{X_i-\mu}{\sigma}$，应用习题 6.10 的结论证明 $W_n\triangleq\sum_{i=1}^{n}Y_i^2$ 的概率密度函数为 n 个自由度的 χ^2，即

$$f_W(W;n)=([(n/2)-1]!2^{n/2})^{-1}w^{(n/2)-1}\exp(-(1/2)w),w>0$$

6.12 $X:N(\mu,\ \sigma^2)$ 的 n 个独立同分布观测为 X_i，令 $\hat{\mu}=n^{-1}\sum_{i=1}^{n}X_i$，$\hat{\sigma}^2=(n-1)^{-1}\sum_{i=1}^{n}(X_i-\hat{\mu})^2$。证明 $\hat{\mu}$ 和 $\hat{\sigma}^2$ 相互独立(提示：可采用矩生成函数，若不成功可参考附录 F)。

6.13 令 $X:N(\mu,\ \sigma^2)$ 及 $W_n:\chi_n^2$ 为两个独立随机变量，令 $Y\triangleq\dfrac{X-\mu}{\sigma}$。

(a) 证明 Y 和 W_n 的联合概率密度为

$$f_{YW_n}(y,w)=\frac{1}{\sqrt{2\pi}}\exp(-0.5y^2)\times\frac{W^{(n-2)/2}\exp(-0.5w)}{[(n-2)/2]!2^{n/2}},-\infty<y<\infty,w>0$$

(b)令 $T \triangleq \frac{Y}{\sqrt{W_n/n}}$，证明 $f_T(t;n) = \frac{[(n-1)/2]!}{\sqrt{n\pi}[(n-2)/2]!} \times \frac{1}{[1+t^2/n]^{(n+1)/2}}$，$-\infty < t < \infty$，即为“学生” t 概率密度函数(提示：应用适当的二元变换)。

6.14 令 W_m 和 V_n 分别为独立 χ_m^2 和 χ_n^2 变量，则 $F \triangleq \frac{W_m/m}{V_n/n}$ 为具有 m 和 n 个自由度的 F 分布，证明 T^2 为具有 1 和 n 个自由度的 F 分布。

6.15 采用 MATLAB，Excel，或其他科学计算程序产生正态随机变量 $X:N(0,1)$ 均值的 95% 置信区间。每个区间计算采用 50 个观测并重复实验 50 次，每次实验记录区间长度和是否包括均值，均值应为零。每个区间采用 100 个观测重复一次。

6.16 证明式(6.2-3)的样本方差的无偏性。

6.17 证明式(6.2-3)的样本方差的一致性。

6.18 证明应用式(6.3-7)对 σ_X 的一次估计的方差为 $\mathrm{Var}(\hat{\sigma}) = (\frac{\pi}{2} - 1)\sigma_X^2 \approx 0.57\sigma_X^2$。

6.19 观测到下列正态数据：

3.810e+0	2.550e+0	−1.150e+0	−1.230e+0	5.640e−1
−1.420e−2	−2.370e+0	2.500e+0	1.130e+0	−1.670e+0
−3.000e+0	1.730e+0	−7.040e−1	−9.680e−1	−9.630e−1

求分布方差 σ_X^2 的 95% 置信区间。

6.20 证明式(6.6-6)的数值 a 为 $a = z_{(1+\delta)/2}$，$-a = z_{(1-\delta)/2}$，即 a 为 $Z:N(0,1)$ 的 $(1+\delta)/2$ 百分点。

6.21 用随机变量发生器得到 20 个 $N(0,2)$ 分布的正态随机变量，应用式(6.4-12)的估计器估计其方差，并应用式(6.4-2)的无偏方差估计器估计同一批数据。

6.22 令 $X:N(\mu_X, \sigma_X^2)$ 的三个观测为 X_1, X_2, X_3，令 $V_i \triangleq \frac{X_i - \hat{\mu}_X(n)}{\sigma_X}$，$i=1,2,3$。证明 $\sum_{i=1}^{3} V_i^2$ 为两个自由度的卡方分布。

6.23 应用正态随机变量发生器得到 20 个满足 $X:N(0,2)$ 分布的数据，计算其方差的 95% 置信区间。试将 5% 的误差等分到卡方分布概率密度函数的两边从而得到近似解答。

6.24 应用正态随机变量发生器得到 20 个满足 $X:N(0,1)$ 分布的数据 $\{x_i, i=1,\cdots,20\}$，重复此过程获得一组新的数据 $\{y_i, i=1,\cdots,20\}$。产生两组数据 $\{w_i, i=1,\cdots,20\}$ 和 $\{v_i, i=1,\cdots,20\}$，满足 $w_i = x_i + (y_i/\sqrt{2})$ 及 $v_i = x_i - (y_i/\sqrt{2})$，则随机变量 W 和 V 具有相关系数 $\rho_{VW} = 1/3$)(证明此结论)，计算样本相关系数 $\hat{\rho}'_{VW}$ 的大小。

6.25 证明式(6.3-1)的方差估计函数是无偏且具有一致性。

6.26 求概率质量函数为 $P_X(n) = (1-a)a^n u(n)$ 的随机变量 X 的均值和方差，并计算 X 均值的 95% 置信区间。

6.27 在例 6.5-3 中，说明了 $P\left[-a \leqslant \frac{\hat{p}-p}{\sqrt{pq/n}} \leqslant n\right] = \delta$ 等价于 $P[(p-\hat{p})^i \leqslant a^2pq/n] = \delta$，试证明之。

6.28 证明均值的最大似然估计具有均方一致性。

6.29 计算指数概率密度函数参数 λ 的最大似然估计，证明似然函数实际上在 MLE 处为最大值。

6.30 计算二项式概率质量函数参数 $p \triangleq P[\text{success}]$ 的最大似然估计。

6.31 计算 $f_X(x) = (b-a)^{-1}(u(x-a) - u(x-b))$ 中参数 $a, b(a<b)$ 的最大似然估计。

6.32 考虑线性模型 $\boldsymbol{Y} = \boldsymbol{Ia} + \boldsymbol{bx} + \boldsymbol{V}$，其中

$$\boldsymbol{Y} \triangleq [Y_1, \cdots, Y_n]^{\mathrm{T}}$$
$$\boldsymbol{V} \triangleq [V_1, \cdots, V_n]^{\mathrm{T}}$$
$$\boldsymbol{I} \triangleq n \times n \text{ 单位矩阵}$$
$$\boldsymbol{x} \triangleq [x_1, \cdots, x_n]^{\mathrm{T}}$$
$$\boldsymbol{a} \triangleq [a, \cdots, a]^{\mathrm{T}}$$

a, b 为待定的常数。假定 $\{V_i, i=1, \cdots, n\}$ 为 n 个独立同分布 $N(0, \sigma^2)$ 随机变量，对每一 $i=1, \cdots, n$, x_i, $i=1, \cdots, n$ 为随 i 变化常数，称为控制变量。

(i)证明 $\{Y_i, i=1, \cdots, n\}$ 服从 $N(a+bx_i, \sigma^2)$。

(ii)写出似然函数并证明当 $\sum_{i=1}^{n}(Y_i-(a+bx_i))^2$ 最小时其最大化。

(iii)证明 a 的最大似然估计为 $\hat{a}_{\mathrm{ML}}=\hat{\mu}_Y-\hat{b}_{\mathrm{ML}}\bar{x}$, b 的最大似然估计为 $\hat{b}_{\mathrm{ML}}=\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})Y_i}{\sum_{i=1}^{n}(x_i-\hat{x})^2}$，其中

$$\hat{\mu}_Y \triangleq (1/n)\sum_{i=1}^{n} Y_i$$
$$\bar{x} \triangleq (1/n)\sum_{i=1}^{n} x_i$$

6.33 假定学院的足球队员体重服从均值220磅、标准差20磅的正态分布，确定随机变量体重的95%百分点。

6.34 计算几何分布随机变量的中值。

6.35 计算卡方随机变量的均值和中值。

6.36 证明30%百分点的估计 $x_{0.3}$ 是由内插公式 $Y_4+\dfrac{(Y_4-Y_3)(0.3-4/11)}{1/11}\sim x_{0.3}$ 给出，其中随机变量 $\{Y_i, i=1, \cdots, 10\}$ 为非排序的独立同分布随机变量 $\{X_i, i=1, \cdots, 10\}$ 排序得出。

6.37 以0.99的概率覆盖50%百分点需要多少样本？提示：应用式 $P[Y_1<x_{0.5}<Y_n]=\sum_{i=1}^{n-1}\binom{n}{i}(1/2)^n\approx 0.99$，其中 $Y_1\triangleq\min(X_1, X_2, \cdots, X_n)$, $Y_n\triangleq\max(X_1, X_2, \cdots, X_n)$。

*6.38 证明排序随机变量 Y_1, Y_n 的联合概率密度函数为

$$f_{Y_1Y_n}(y_1,y_n)=n(n-1)\left(F_X(y_n)-F_X(y_1)\right)^{n-2}f_X(y_1)f_X(y_n),-\infty<y_1<y_n<\infty$$

其中 $Y_1\triangleq\min(X_1, X_2, \cdots, X_n)$, $Y_n\triangleq\max(X_1, X_2, \cdots, X_n)$。提示：考虑 $Y_1, Y_2, \cdots, Y_n$ 的联合概率密度函数并对首尾除外部分进行积分。

*6.39 $\{Y_i, i=1, \cdots, n\}$ 为一排序随机变量集，定义集合范围 R 为 $R\triangleq Y_n-Y_i$。考虑概率密度函数为 $f_X(x)=u(x)-u(x-1)$ 的 X 的6个观测 $\{X_i, i=1, \cdots, 6\}$，其中 $u(x)$ 为单位阶跃函数。证明 $f_R(r)=30r^4(1-r)$, $0<r<1$。提示：应用结论 $f_{Y_1Y_n}(y_1, y_n)=n(n-1)(F_X(y_n)-F_X(y_1))^{n-2}f_X(y_1)f_X(y_n)$, $-\infty<y_1<y_n<\infty$，并定义两个随机变量 $R\triangleq Y_n-Y_1$, $S\triangleq Y_1$，求出 $f_{RS}(r, s)$ 后对 S 求积分可得的结果。

参考文献

6-1 A. M. Mood and F. A. Graybill, *Introduction to the Theory of Statistics*, 2nd Edition. New York: McGraw-Hill, 1963.

6-2 A. Papoulis, *Probability &Statistics*. Englewood Cliffs, NJ. Prentice Hall, 1990.

6-3 H. Stark and J. G. Brankov, "Estimating the Standard Deviation from Extreme Gaussian Values," *IEEE Signal Processing Letters*, Vol. 11, No. 3, 2004, pp. 320-322.

6-4 W. Luo, "A Comment on 'Estimating the Standard Deviation from Extreme Gaussian Values'," *IEEE Signal Processing Letters*, Vol. 12, No. 2, 2005, p. 109.

6-5 K. S. Miller, *Multidimensional Gaussian Distributions*. New York: John Wiley, 1964.

6-6 J. N. Franklin, *Matrix Theory*. Upper Saddle River, NJ. Prentice Hall, 1968.

6-7 K. Fukunaga, *Introduction to Statistical Pattern Recognition*. 2nd edition. New York: Academic, 1990.

6-8 M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 2, 3rd Edition. London, England: Charles Griffin and Co., 1951.

6-9 T. O. Lewis and P. L. Odell, *Estimation in Linear Models*, Upper Saddle River, NJ. Prentice-Hall, 1971.

补充文献

A. L. Garcia, *Probability, Statistics and Random Processes for Electrical Engineering*, 3rd Edition, Upper Saddle River, NJ. Prentice Hall, 2008.

S. M. Kay, *Fundamentals of Signal Processing*, Upper Saddle River, NJ. Prentice Hall, 1993.

G. Straud, *Linear Algebra and its Applications*, 3rd Edition. New York: Saunders, 1900.

(Online) Y. Cohen and J. Y Cohen, Chapter 9 in *Statistics and data with R: An Applied Approach through Examples*. New York: Wiley Online Library, 2008.

J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions of the Royal Society of London*, *Series A*, Vol. 231, 1933, pp. 289-337.

T. A. Severini, "On the Relationship Between Bayesian Interval Estimates," *Journal of the Royal Statistical Society*, *Series B*, Vol. 53, No. 3, 1991, pp. 611-618.

J. E. Bernardo and A. Smith, *Bayesian Theory*. New York: Wiley, p. 259.

(Online) Tutor Vista. com: Probability Calculator

(Online) Statistics Help@ Talk Stats Forum

第7章 假设检验

假设检验在统计判决理论的广泛应用中是一个重要的主题，统计判决理论还包括其他方面，比如预测、统计回归、游戏理论、统计建模及信号处理中很多基础部分，然而，假设检验通常是这些更加高级应用领域的基础①。

假设的形式如下：我们做出一个假设，即一个参数为一个确定值，或于某一个确定范围内，或一个确定事件发生，则对应备选假设②参数为其他数值，或于其他范围，或一个事件没有发生。然后，基于实际数据，我接受（或拒绝）此假设或接受（拒绝）备选假设。参数估计和假设检验很明显是相关的，比如，判决一集合均值等于另一集合的已知均值假设成立。实质上等价于对前一集合的均值做出估计，并判定其与已知均值足够接近，以至于认为是相等的。

在现实中，我们经常在没有掌握所有事实，或我们的知识来自于固有概率的观测情况下被迫做出判决。我们都（大概）知道某重度吸烟者在80岁时过得很好，同时，也知道某非吸烟者在50岁时死于肺癌，这会意味着吸烟与肺癌无关？以前，烟草公司负责人会说是，而癌症专家会说不。但根据后来的证据，现在已没有一个理智的人会否认吸烟是会增加肺癌死亡率的。尽管如此，这不同于一个人从20楼摔到水泥地面，重度吸烟者并不总是死于肺癌或相关疾病，吸烟与肺癌的关系本质上是概率的。下面的各部分我们将讨论在概率环境下做出判决的方法。

7.1 贝叶斯判决理论

在随机环境下采用贝叶斯方法做出判决可能是人类想出的最合理的步骤。遗憾的是，采用基本形式的贝叶斯需要我们不能精确获得或不可能获得的信息。我们用下述例子阐述贝叶斯理论的应用及其存在的不足之处。

例7.1-1 （判决是否做手术）假定你是一个外科医生，你的病人X射线显示左肺有个瘤子，病人40岁，没有吸烟史，其他方面很健康。让我们将问题简化为两种可能状态：(1)瘤子是早期癌细胞，不手术将扩散并致死；(2)瘤子是良性的，不会增加健康风险。将前者表示为ζ_1，后者表示为ζ_2。读者会发现输出空间（看“样本空间”部分）Ω仅有两个点，即$\Omega=\{\zeta_1,\zeta_2\}$，当然在复杂情况下会有更多。外科医生的工作是做出对病人最有利的判决（并做出相应治疗）。问题是医生不通过手术无从知道到底是ζ_1还是ζ_2，这有两个选项：手术(a_1)或不手术(a_2)。

哪种方式最佳通常是不明显的，然而，在其他情况不变时，这个案例的合理最佳是判决将使得病人正常寿命缩小年限最小化。有四种情形需要考虑：

(1) 医生判决瘤子为良性的，不做手术

① 有几本教科书与此相关，如参考文献[7-1]至参考文献[7-4]。

② 在备选假设也通常简称为备选，因此，可能会看到有“我们考察假设…对备选…”。

(2) 医生判决瘤子是癌细胞，不做手术

(3) 医生判决瘤子为良性的，做手术

(3) 医生判决瘤子是癌细胞，做手术

先验数据表明中年非吸烟者肺部的瘤子70%是良性的，因此，瘤子为癌细胞的概率是30%，同时医生也有表7.1-1的数据。

表达式$\{l(a_i, \zeta_j), i=1, 2; j=1, 2\}$称为代价函数，$l(a_i, \zeta_j)$为当状态是$\zeta_j$而做出$a_i$行为所付出的代价。读者可能会问为什么$l(a_1, \zeta_1)=l(a_1, \zeta_2)=5$而不是零，医生知道手术是有风险的，即便是健康的人也会受到术后MRSA① 或革兰氏阴性菌的感染②，除非绝对必要，大多数医生会选择非损伤性手术而回避损伤性手术。因此，由于感染或医生承担的其他风险，加上病人在术中死亡等因素，我们分配的平均寿命损失年限为5年。

表7.1-1

判　决	实际状态	自然寿命减少的年限
不手术(a_2行为)	良性组织(ζ_2)	$l(a_2,\zeta_2)=0$
不手术(a_2行为)	癌细胞(ζ_1)	$l(a_2,\zeta_1)=35$
手术(a_1行为)	良性组织(ζ_2)	$l(a_1,\zeta_2)=5$
手术(a_1行为)	癌细胞(ζ_1)	$l(a_1,\zeta_1)=5$

下面，介绍判决函数d的概念。判决函数d是观测数据的函数，因此写为$d(X_1, X_2, \cdots, X_n)$，其中$\{X_i, i=1, \cdots, n\}$是对随机变量(RV)X的n个独立同分布的观测。判决函数$d(X_1, X_2, \cdots, X_n)$帮助指导医生采取哪种行为，即a_1或a_2。在这个例子中，我们限定为单一观测X，即瘤子边界长度的平方与区域的比值，这是一个瘤子边缘不规则形的测度：边缘越不规则，瘤子为癌细胞的可能性就越大(参见图7.1-1)。因此，我们预计癌细胞(ζ_1)的随机变量X相比于良性时(ζ_2)的X在大多数情况下会得到较大值。此时得到X的实现即为数据集。

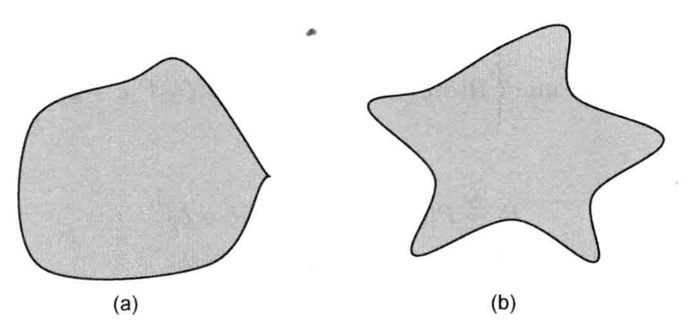


图7.1-1 (a)良性瘤子有规则的边界；(b)癌细胞有不规则边界

令$f(x;\zeta_1)$和$f(x;\zeta_2)$分别表示X在ζ_1和ζ_2时的概率密度函数(参见图7.1-2)，为简化及突出可视效果，在本例中假定概率密度函数为连续的单峰形式，另外，如图7.1-2所示，假定存在一常数c，当数据落在c的右侧，则被视为有癌细胞的证据，若数据落在c的左侧，则视为癌细胞不存在。若有证据表明癌细胞存在，则将执行a_1行为，否则执行a_2行为。既然这

① 耐甲氧西林金黄色葡萄球菌。

② 此种病菌在医院流行并且很难处理。

是概率事件，一定有错误发生，因此

$$P[a_1|\zeta_2] = \int_c^{\infty} f(x;\zeta_2)\mathrm{d}x \tag{7.1-1}$$

为证据显示有癌细胞而需要手术，但事实不是癌细胞的错误概率。同样

$$P[a_2|\zeta_1] = \int_{-\infty}^{c} f(x;\zeta_1)\mathrm{d}x \tag{7.1-2}$$

为证据显示不是癌细胞而不需要手术，但事实是癌细胞的错误概率。

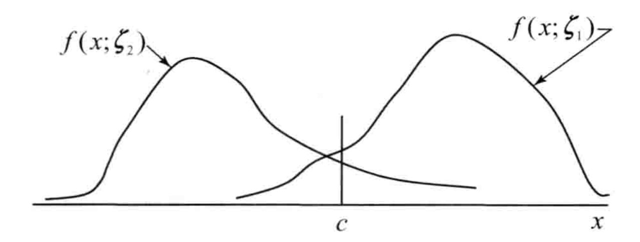


图7.1-2 存在 c 值(待确定)使得期望风险最小。在区域 $\Gamma_2 \triangleq (-\infty, c]$ 内的数据更可能与良性结论并导致 a_2 行为(不手术)相关,而区域 $\Gamma_1 \triangleq [c,\infty)$ 内的数据更可能与癌细胞结论并导致 a_1 行为(手术)相关

当状态为 ζ 而判决规则为 d 的代价的条件期望称为风险 $R(d,\zeta)$，因此

$$R(d;\zeta_2) = l(a_1;\zeta_2)P[a_1|\zeta_2] + l(a_2;\zeta_2)P[a_2|\zeta_2] \tag{7.1-3}$$

和

$$R(d;\zeta_1) = l(a_1;\zeta_1)P[a_1|\zeta_1] + l(a_2;\zeta_1)P[a_2|\zeta_1]$$

最后，用 $B(d)$ 表示的期望风险定义为①

$$B(d) = R(d;\zeta_1)P[\zeta=\zeta_1] + R(d;\zeta_2)P[\zeta=\zeta_2] \tag{7.1-4}$$

其将被最小化。通过最小化 $B(d)$ 得到判决函数 d^* 即为贝叶斯准则。

因此

$$B(d^*) = \min_d \left\{ R(d;\zeta_1)P[\zeta=\zeta_1] + R(d;\zeta_2)P[\zeta=\zeta_2] \right\} \tag{7.1-5}$$

概率

$$P_1 \triangleq P[\zeta=\zeta_1], P_2 \triangleq P[\zeta=\zeta_2]$$

称为状态的先验概率。根据前面引入的符号，$B(d)$ 可以写为

$$B(d) = P_1 \times l(a_2,\zeta_1) + P_2 \times l(a_2,\zeta_2)$$
$$+ \int_c^{\infty} \left\{ P_2 f(x;\zeta_2)\left[l(a_1,\zeta_2) - l(a_2,\zeta_2)\right] - P_1 f(x;\zeta_1)\left[l(a_2,\zeta_1) - l(a_1,\zeta_1)\right] \right\}\mathrm{d}x \tag{7.1-6}$$

其中我们选择 c 使得 $B(d)$ 最小化。若 $B(d)$ 表达式中的积分为正，则使 $B(d)$ 增加，若积分为负，则使 $B(d)$ 减小。实际上，若选择 c，比如 $c=c^*$，从而使得 (c^*, ∞) 包含所有积分为负的

① 符号 B 由令人尊敬的数学家/哲学家托马斯·贝叶斯采用(1702–1761)。

值，而所有积分为正的值位于其外，则就得到最小化 $B(d)$。使积分为负的结果[①]（参见“事件”部分）表示为

$$\frac{f(X;\zeta_1)}{f(X;\zeta_2)} > \frac{[l(a_1,\zeta_2)-l(a_2,\zeta_2)]P_2}{[l(a_2,\zeta_1)-l(a_1,\zeta_1)]P_1} \triangleq k_b \tag{7.1-7}$$

即为贝叶斯判决规则，这说明所有(c^*, ∞)的结果导致 a_1 行为（手术）。同样对所有$(-\infty, c^*)$的结果，即满足

$$\frac{f(X;\zeta_1)}{f(X;\zeta_2)} < k_b$$

导致 a_2 行为（不手术）。常数 c 即满足下式的值。

$$\frac{f(c^*;\zeta_1)}{f(c^*;\zeta_2)} = k_b$$

此例中先验概率是从数千例相似病例治疗中积累的信息获得的，一个 40 岁、非吸烟男性的瘤子是良性的可能性大于癌细胞的可能性。举例来说，它可能是一个无害的残留的疤痕组织，或者是看似一个瘤子的血管交叉点。为简化起见，我们假定其概率分别为 $P_1=0.7$，$P_2=0.3$，则根据式(7.1-6)得到

$$B(d) = 10.5 + \int_c^\infty (3.5f(x;\zeta_2) - 9f(x;\zeta_1))\,\mathrm{d}x$$

这表明常数 c^* 满足 $3.5f(c^*;\zeta_2) - 9f(c^*;\zeta_1) = 0$。可得贝叶斯判决规则为

$$f(X;\zeta_1)/f(X;\zeta_2) > 0.39 \rightarrow \text{手术}$$
$$f(X;\zeta_1)/f(X;\zeta_2) < 0.39 \rightarrow \text{不手术}$$

在例 7.1-1 中，做判决仅仅依据单一随机变量。然而在很多问题中，会基于许多独立同分布随机变量做出判决。在这种情况下，贝叶斯判决规则的形式为

$$\begin{aligned}
&\frac{f(X_1;\zeta_1)\cdots f(X_n;\zeta_1)}{f(X_1;\zeta_2)\cdots f(X_n;\zeta_2)} > \frac{[l(a_1,\zeta_2)-l(a_2,\zeta_2)]P_2}{[l(a_2,\zeta_1)-l(a_1,\zeta_1)]P_1} \triangleq k_b,\ \text{接受}\ \zeta_1\ \text{状态} \\
&\frac{f(X_1;\zeta_1)\cdots f(X_n;\zeta_1)}{f(X_1;\zeta_2)\cdots f(X_n;\zeta_2)} < \frac{[l(a_1,\zeta_2)-l(a_2,\zeta_2)]P_2}{[l(a_2,\zeta_1)-l(a_1,\zeta_1)]P_1} \triangleq k_b,\ \text{拒绝}\ \zeta_1\ \text{状态}
\end{aligned} \tag{7.1-8}$$

读者需认识到式(7.1-8)的分子和分母均为在第 6 章中讨论的似然函数 $L(\zeta_j) = \prod_{i=1}^{n} f_X(X_i, \zeta_j)$，$j=1, 2$，即式(7.1-8)为两个似然函数的比值（似然比）与常数相比较，这通常称为似然比检验（LRT）。式(7.1-8)中的常数 k_b 称为贝叶斯门限。

各个贝叶斯准则都导致 LRT，但不是所有 LRT 都是贝叶斯准则的结果。贝叶斯准则寻求平均风险的最小化，其他似然比形式的检验可能遵循其他准则，比如说，对于给定错误概率，使得似然比最大。实现贝叶斯准则的一个问题是先验概率 P_1 和 P_2 通常是未知的，另一个问题是对于特定行为指定一个合理的代价可能有困难。举例来说，你准备敲破一打鸡蛋来做一个煎蛋饼，你准备采用贝叶斯准则使得代价最小，即工作量最小。你将选用一个或两个碗，此处的随机因素是鸡蛋是否是好的，假定每一百个鸡蛋中平均有一个坏的，若你用一个碗，而在你认识到坏鸡蛋前已经打入其他鸡蛋中，则混合的鸡蛋会浪费。若采用两个碗，在混入其他鸡

① 回想对（实）随机变量 X 映射中，事件是在实轴的间隔点。

蛋前用一个小碗来观察新打出的鸡蛋的好坏，用一个大碗来装所有敲开的好鸡蛋，那么就会避免因一个坏鸡蛋浪费好鸡蛋。然而，当这些完成后，不得不洗两个碗而不是一个。如何合理定义此例中的代价？尽管这个例子不是非常严肃，它还是说明了贝叶斯准则应用中的问题。另一个问题是对极少发生事件的先验概率估计有困难，举例来说，假定一个国家想用它的反导弹武器来防御敌对邻国的进攻，如果防御战略是根据贝叶斯准则来设计的，需要知道敌国进攻的先验概率，这如何通过合理的方式来估计？

7.2 似然比检验

因为先验概率经常未知，因此代价函数不容易定义，我们放弃最小化期望风险的约束，修改贝叶斯判决准则为

$$\begin{aligned}&\frac{f(X_1;\zeta_1)\cdots f(X_n;\zeta_1)}{f(X_1;\zeta_2),\ldots f(X_n;\zeta_2)}>k,\ \text{接受}\zeta_1\text{状态}\\&\frac{f(X_1;\zeta_1)\cdots f(X_n;\zeta_1)}{f(X_1;\zeta_2),\ldots f(X_n;\zeta_2)}<k,\ \text{拒绝}\zeta_1\text{状态}\end{aligned}\tag{7.2-1}$$

其中门限值 k 是由贝叶斯以外的准则确定的。一般准则都与当陈述是真的而拒绝的概率，或是当对立陈述是真的而接受的概率相关，这类检验称为似然比检验，它检验一个简单假设(陈述)与一个简单备选(对立陈述)。为简单起见，我们定义似然比检验随机变量为

$$\begin{aligned}\Lambda&\triangleq f(X_1;\zeta_1)\cdots f(X_n;\zeta_1)/f(X_1;\zeta_2)\cdots f(X_n;\zeta_2)\\&=L(\zeta_1)/L(\zeta_2)\end{aligned}\tag{7.2-2}$$

并通过一个例子说明其应用。

例 7.2-1 (检验一个食品的声明)考虑一个健康食品制造商声明其开发了一个小吃，可以为肥胖小孩减轻体重①。据称美味的小吃店可以抑制胃口，因此降低在就餐时间之间对薯条、汉堡、苏打饼、巧克力等肥胖食品的欲望。为检验此声明的有效性，我们选取 n 个小孩(一群小孩中选取的一个子集)，提供他们减肥小吃。一个月后，小孩的平均体重为 98 磅，标准差为 5 磅，这群小孩中的其他小孩，即没有提供减肥小吃的小孩，平均体重 102 磅，标准差为 5 磅。我们做出减肥小吃对控制体重无效的假设②，称为零假设，用 H_1 表示，备选假设用 H_2 表示，即减肥小吃对控制体重有效③。哪个假设指定为 H_1 并不重要，但是一旦做出选择，在整个问题中需要保持一致。在没有清晰的代价函数情况下，我们关注其错误概率，定义

$$\alpha\triangleq P[\text{当 }H_1\text{为真时判决 }H_2\text{为真}]$$

$$\beta\triangleq P[\text{当 }H_2\text{为真时判决 }H_1\text{为真}]$$

用独立同分布随机变量 X_i，$i=1$，…，n 表示 n 个小孩体重，假定两群小孩体重是以均值为中心的正态分布④，即

---

① 小孩肥胖是美国一个严重的问题，从目前热量消费速度推断，到 2020 年，将有四分之三的美国人超重或肥胖(消费者报告，2010 年 12 月，第 11 页)。

② 这句话的意思是：这是一个暂时接受的假设，没有证据明显支持。若后续信息不支持，这个假定会被拒绝。

③ 在许多书中，零假设用 H_0 表示，备选假设用 H_a 表示，我们选用了数字下标表示法。

④ 此处正态特性只在概率密度函数均值附近 σ 取值较小时有效，显然在尾部是无效的，例如，如何能取一个“负体重”呢？

$$f(x_i, H_1) = \frac{1}{5\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i - 102}{5}\right)^2\right]$$
$$f(x_i, H_2) = \frac{1}{5\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_i - 98}{5}\right)^2\right]$$

注意到$f_X(x_i, H_1)(f_X(x_i, H_2))$分别是$X_i$, $i=1, \cdots, n$在$H_1(H_2)$假设下的概率密度函数。

根据式(7.2-2)有

$$\Lambda = \exp\left(\frac{1}{2}\sum_{i=1}^{n}\left[\left(\frac{X_i - 98}{5}\right)^2 - \left(\frac{X_i - 102}{5}\right)^2\right]\right)$$

可化简为

$$\Lambda = K_n \exp\left(\frac{4n}{25}\hat{\mu}_X(n)\right)$$

其中$\hat{\mu}_X(n) \triangleq n^{-1}\sum_{i=1}^{n} X_i$，$K_n$常数，与样本数$n$有关，与$\{X_i, i=1, \cdots, n\}$无关。判决函数为

$$\text{若 } K_n \exp\left(\frac{4n}{25}\hat{\mu}_X(n)\right) > k_n,\quad H_1 \text{ 为真},(H_2 \text{ 为假})$$
$$\text{若 } K_n \exp\left(\frac{4n}{25}\hat{\mu}_X(n)\right) < k_n,\quad H_2 \text{ 为真},(H_1 \text{ 为假})$$

由于Λ的自然对数($\ln\Lambda$)是Λ的增函数(参见图7.2-1)，将不同常量简化为单个常量后，可以用自然对数形式表示判决函数为

$$\text{若 } \hat{\mu}_X(n) > c_n,\quad H_1 \text{ 为真}$$
$$\text{若 } \hat{\mu}_X(n) < c_n,\quad H_2 \text{ 为真}$$

其中c_n是与测试小孩数目n相关的常数，并由我们选定的准则确定。若H_1为真，则$\hat{\mu}_X(n)$分布为$N(102, 25/n)$，即

$$f_{\hat{\mu}}(x, H_1) = \frac{1}{\sqrt{50\pi/n}} \exp\left(-\frac{1}{2}\left[\frac{x - 102}{5/\sqrt{n}}\right]^2\right)$$

若H_2为真，则$\hat{\mu}_X(n)$分布为$N(98, 25/n)$，即

$$f_{\hat{\mu}}(x, H_2) = \frac{1}{\sqrt{50\pi/n}} \exp\left(-\frac{1}{2}\left[\frac{x - 98}{5/\sqrt{n}}\right]^2\right)$$

$\hat{\mu}_X(n)$在H_1和H_2情况下的概率密度函数如图7.2-2所示。

假定准则中我们指定$\alpha=0.025$，回想$\alpha \triangleq P$[当H_1为真时判决H_2为真]，则有

$$0.025 = \int_{-\infty}^{c_n} f_{\hat{\mu}}(x, H_1)\,\mathrm{d}x = F_{\hat{\mu}}(c_n) = F_{\mathrm{SN}}\left(\frac{c_n - 102}{5/\sqrt{n}}\right) = F_{\mathrm{SN}}(z_{0.025})$$

根据正态表并简化，给出门限值为$c_n = 102 - (9.8/\sqrt{n})$。与其他地方一样，符号$F_{\mathrm{SN}}(z)$表示标准正态随机变量在$z$处的累积密度函数，因此$H_1$为真判决需满足$(102-(98/\sqrt{n}) < \hat{\mu}_X(n) < \infty)$，既然实轴上间隔即为随机变量的映射，也可以写成$(102-(9.8/\sqrt{n}), \infty)$。样本数$n$对门限的影响在图7.2-3中显示，检验的势随样本数增加而增加，如图7.2-4所示。势的增加意味着当H_2为真时错误判决概率下降。

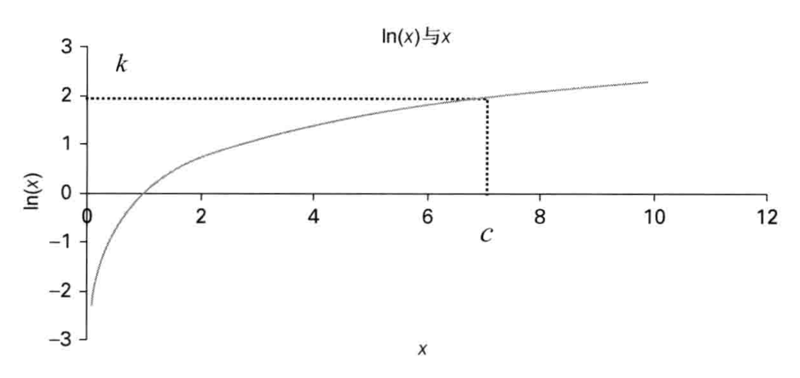


图 7.2-1 x 的自然对数是 x 的增函数

错误概率 α 称为第一类错误概率或检验显著性水平①，概率 $P \triangleq 1-\beta$ 称为检验的势，β 称为第二类错误概率。检验的势就是当备选假设为真时，我们拒绝零假设的概率。尽管都是错误概率，一般来说，不可能使得 α 和 β 同时很小，通常有 $\alpha+\beta=1$。

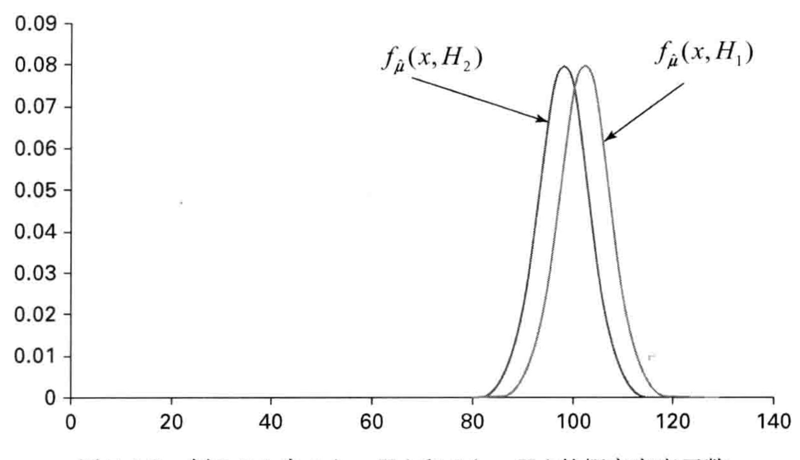


图 7.2-2 例 7.2-1 中 $f_{\hat{\mu}}(x, H_1)$ 和 $f_{\hat{\mu}}(x, H_2)$ 的概率密度函数

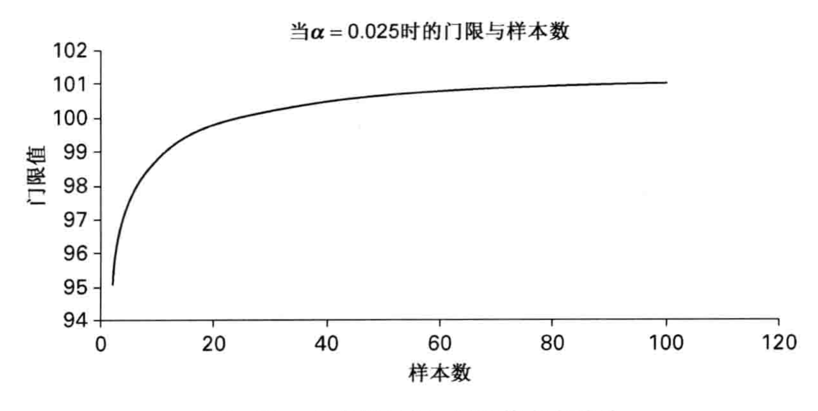


图 7.2-3 样本数增加，门限值向右移动

① 错误概率 α 有时称为检验的尺寸。

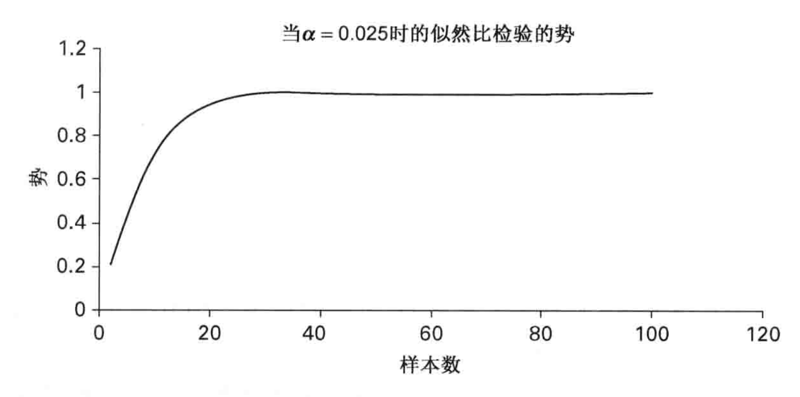


图 7.2-4　检验的势随样本数增加而增加，这是好现象。最佳检验是对于给定n和α对势的最大化，此例中，检验实际上是最佳检验

可能会有读者对例 7.2-1 中的问题提出疑问，也就是说，既然吃减肥小吃的小孩体重比其他小孩平均轻 4 磅，为何不根据这个证据判定减肥小吃对体重控制的有效性？这会忽略即使对于体重较大的一群小孩，98 磅也在 102 磅的一个标准差内的事实，这意味着如果样本数较小，我们可能会得出小吃有效的错误结论。此外，这个简单方法不会告诉我们误判的概率。

例 7.2-2　（正态分布人群均值的差异）营养学界有一种认识，即把援助第三世界营养不良的儿童通过饮食中富含 w-3 脂肪酸（比如鱼油）和复合碳水化合物（全麦、玉米、糙米等）的方法，在提高健康的同时，能在 13 岁时提高 IQ 值 10 个点。为测试这个结论，需要测量吃这种饮食长大的小孩的 IQ，并与吃当地食物长大的小孩 IQ 进行对比，通常数据是 n 个吃实验食物的小孩 IQ 的样本均值$\hat{\mu}(n)$。如果我们将真实但未知的均值表示为μ，检验形式为：$H_1:\mu_{\text{IQ}}=110$ 对 $H_2:\mu_{\text{IQ}}=100$，有几个等效的检验，例如，$H_1:\mu=a$ 对 $H_2:\mu\neq a$，及 $H_1:a<\mu<b$ 对 $H_2:\mu<a$，$\mu>b$。我们对于方差为 σ^2 的正态分布人群采用基本检验 $H_1:\mu=b$ 对 $H_2:\mu=a(b>a)$，假定随机样本数为 n，即有 n 个独立同分布随机变量 X_1，…，X_n，则若 H_1 为真，有 X_i: $N(b,\sigma^2)$，若 H_2 为真，有 X_i:$N(a,\sigma^2)$，似然比检验随机变量为

$$\Lambda=\frac{\prod\limits_{i=1}^{n}\left(2\pi\sigma^2\right)^{-1/2}\exp\left(-\frac{1}{2}\left[\frac{X_i-b}{\sigma}\right]^2\right)}{\prod\limits_{i=1}^{n}\left(2\pi\sigma^2\right)^{-1/2}\exp\left(-\frac{1}{2}\left[\frac{X_i-a}{\sigma}\right]^2\right)}\tag{7.2-3}$$

化简后取对数，常量合并可得检验为

$$\begin{aligned}&\text{若 }\hat{\mu}(n)>c_n,\ H_1\text{ 为真},(H_2\text{ 为假})\\&\text{若 }\hat{\mu}(n)<c_n,\ H_2\text{ 为真},(H_1\text{ 为假})\end{aligned}$$

常量 c_n 由 α 确定，若 $\alpha=0.025$，即有

$$\begin{aligned}\alpha&=P[H_2\text{为真}|H_1\text{为真}]=0.025\\&=\int_{-\infty}^{c_n}\frac{1}{(2\pi)^{0.5}\sigma/\sqrt{n}}\exp\left(-1/2\left[\frac{y-b}{\sigma/\sqrt{n}}\right]^2\right)\mathrm{d}y\\&=\int_{-\infty}^{c_n'}1/(2\pi)^{0.5}\exp\left(-1/2y^2\right)\mathrm{d}y=F_{\text{SN}}\left(\frac{c_n-b}{\sigma/\sqrt{n}}\right)=F_{\text{SN}}(z_{0.025})\end{aligned}$$

其中 $c_n'\triangleq(c_n-b)\sqrt{n}/\sigma$。从标准正态累积密度函数表中，可得 $z_{0.025}=-1.96$，可求得 $c_n=b-$

$(1.96\sigma/\sqrt{n})$，注意到本例与例 7.2-1 的相似之处。检验的势为

$$\begin{aligned} P &= 1 - P[\text{接受}H_1|H_2\text{ 为真}] \\ &= 1-(2\pi\sigma^2/n)^{-1/2}\int_{c_n}^{\infty}\exp\left(-0.5\left[\frac{z-a}{\sigma/\sqrt{n}}\right]^2\right)\mathrm{d}z \\ &= F_{\mathrm{SN}}\left(\frac{b-a-(1.96\sigma/\sqrt{n})}{\sigma/\sqrt{n}}\right) \end{aligned} \tag{7.2-4}$$

读者应认识到，检验的势就是 H_2 为真时判决 H_2 为真的概率。回到引起讨论的 IQ 问题，可知对于 $\alpha=0.025$，$b=110$，$a=100$，$\sigma=10$ 和 $n=25$，判决 H_1 的区域为 $c_n=106.1$ 的右侧区域，换句话说，当事件 $\{106.1<\hat{\mu}(n)<\infty\}$ 发生，说明一份好的食物可以克服营养不良儿童的 IQ 下降，检验的势为 0.999。

纽曼-皮尔逊定理 假定做一个简单假设与简单备选检验，给定 α 情况下使 β 最小，即最大化势 $P=1-\beta$，称为最大势检验。如何实现检验？纽曼-皮尔逊定理(此处无证明)给出了回答。

定理 7.2-1 用 R_k 表示判决域的点集(例如，判决 H_1 为假的区域)，检验的显著性表示为 α，即 $P[\text{接受 } H_2 \mid H_1\text{为真}]\leqslant\alpha$，则对于确定 R_k 的固定值 k，若满足下式，R_k 最大化检验的势 $P\triangleq 1-\beta$

$$\Lambda \triangleq \frac{f(X_1,\zeta_1)\cdots f(X_n;\zeta_1)}{f(X_1,\zeta_2)\cdots f(X_n;\zeta_2)} < k \tag{7.2-5}$$

讨论 纽曼-皮尔逊理论(NPT)在显著性 α 约束情况下的似然比检验为最大势检验，可以理解为最佳检验。定理没有明确给出 R_k，k 和 α 的关系，但在求解问题时是很清晰的。

例 7.2-3 (可得到大鸡蛋的鸡饲料)一个鸡饲料生产商声明，给鸡喂食一个新产品“鸡生长”，可得到比喂食普通饲料更大的鸡蛋。喂食普通饲料的每个鸡蛋平均质量为 60 g，标准差为 4 g，25 只鸡通过喂食“鸡生长”得到平均质量为 62 g 的鸡蛋，标准差为 4 g。假定 $H_1:\mu=\mu_1=62$ 且备选为 $H_2:\mu=\mu_1=60$，检验显著性水平为 0.05，根据 NPT，检验为

$$\Lambda = \frac{\prod_{i=1}^{n}(2\pi 16)^{-1/2}\exp\left(-\frac{1}{2}\left[\frac{X_i-62}{4}\right]^2\right)}{\prod_{i=1}^{n}(2\pi 16)^{-1/2}\exp\left(-\frac{1}{2}\left[\frac{X_i-60}{4}\right]^2\right)} < k$$

判决域 R_k 可得最大势检验，则 $\Lambda=\exp\left(\frac{n\hat{\mu}}{8}+(60)^2-(62)^2\right)$，取对数、常量合并、化简可得检验为

$$\begin{aligned} &\text{若 } \hat{\mu}<c_n,\ H_1\text{ 为假},\ H_2\text{ 为真} \\ &\text{若 } \hat{\mu}>c_n,\ H_1\text{ 为真},\ H_2\text{ 为假} \end{aligned}$$

其中 c_n 为未知常量。为获得 c_n 及判决域 R_k，求解

$$0.05 = \int_{-\infty}^{c_n}\frac{1}{\sqrt{2\pi}(4/5)}\exp\left(-\frac{1}{2}\left(\frac{z-62)}{0.8}\right)^2\right)\mathrm{d}z$$

可得 $c_n=60.7$，$R_k=(0,60.7)$。因此，若 $\hat{\mu}<60.7$，H_1 为假，H_2 为真，检验为最大势且 $P\approx0.81$。

7.3　复合假设

前面的章节中我们提到有这样的检验形式 $H_1: a<\mu<b$ 对 $H_2: \mu<a,\ \mu>b$ 及其他一些形式。这些检验有一个共同点，H_1或H_2之一或两者事件对应的样本空间有多个结果，而在简单假设和简单备选情形中，样本空间只有 ζ_1 和 ζ_2 两个点。在复合假设情形中，检验

$$\Lambda \triangleq \frac{f(X_1,\zeta_1)\cdots f(X_n;\zeta_1)}{f(X_1,\zeta_2)\cdots f(X_n;\zeta_2)}<k \tag{7.3-1}$$

是没有意义的，因为存在多个 ζ 而不仅仅是 ζ_1 和 ζ_2。为理解本节内容，读者需要回想参数估计的最大似然方法中，其思路是使得观测结果最可能得到，从而在似然函数 $L(\theta)$ 中计算得到参数 θ，通常也可通过微分求解但并不一定。到目前为止讨论的问题中，还不需要通过最大化似然函数来求解 θ，因为 θ 仅仅是 ζ_1 和 ζ_2 中的一个。假定感兴趣的参数是均值，也就是 $\theta=\mu$，则在下述问题如 $H_1: \mu=\mu_n$ 对 $H_2: \mu\neq\mu_0$，最大化与 H_2 相关的似然函数需要在参量空间 $(-\infty,\infty)$ 搜索最佳值 μ，也就是说，尽管假设是简单的，但备选不是简单的，称为复合假设。

幸运的是，不是所有复合假设问题都需要进行搜索，我们依然可以采用纽曼-皮尔逊准则及其优良的最大势性质，通过下例说明。

例 7.3-1　($H_1: \mu=\mu_1$ 对备选 $H_2: \mu<\mu_1$ 的假设检验）假定一均值为 μ 方差为 σ^2 的正态集，第一眼看来似乎 $H_2: \mu<\mu_1$ 的似然函数需要进行搜索，然而，可以将 H_2 做一轻微调整从而得到简单假设对简单备选形式。即修正问题为 $H_1: \mu=\mu_1$ 对 $H_2': \mu=\mu_2<\mu_1$，其中 μ_2 为任意数，则

$$\Lambda=\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(X_i-\mu_1)^2-\sum_{i=1}^{n}(X_i-\mu_2)^2\right)\right)<k \tag{7.3-2}$$

为 H_1 判决域的似然比检验。化简、取对数并常数合并后可得检验：若 $\hat{\mu}<c_n$，H_1为假。为计算常量 c_n，根据前述方法，即第一类错误准则，也就是检验的显著性水平求解，因此，若 $\alpha=0.01$，概率密度函数 $f_{\hat{\mu}}(z;\mu_1)=N(\mu_1,\sigma^2/n)$，求解

$$0.01=\frac{\sqrt{n}}{\sqrt{2\pi}}\int_{-\infty}^{c_n}\exp\left[-0.5\left(\frac{z-\mu_1}{\sigma/\sqrt{n}}\right)^2\right]\mathrm{d}z$$

可得 $c_n=\mu_1-2.32\sigma/\sqrt{n}$。这样，若 $\hat{\mu}<\mu_1-2.32\sigma/\sqrt{n}$，$H_1$ 为假。注意到我们没有计算 μ_2 的真实值。

广义似然比检验（GLRT）

广义似然比检验在求解复合假设问题中很有用。首先，一些似然函数有一个参量，一些似然函数有两个参量，等等，例如，n 个独立同分布指数随机变量的似然函数为 $L(\lambda)=\lambda^n\exp(-\lambda\sum_{i=1}^{n}X_i)u(X_i)$ ①，它是参量 $\theta=\lambda$ 的函数，而 n 个独立同分布正态随机变量的似然函数为

① 函数 $u(x)$ 为单位阶跃：$u(x)=1$，$x\geqslant 0$，其他处为零。

$$L(\mu,\sigma)=\left(2\pi\sigma^2\right)^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left[\frac{X_i-\mu}{\sigma}\right]^2\right)$$

它是两个参量 $\boldsymbol{\theta}=(\mu,\ \sigma)$ 的函数。二维(多维)正态的似然函数是5个参量的函数，即 μ_1，μ_2，σ_1，σ_2，ρ_{12}。我们用符号 $L(\boldsymbol{\theta})$ 表示参量为 $\boldsymbol{\theta}=(\theta_1,\ \theta_2,\ \cdots,\ \theta_k)$ 的似然函数。考虑下述问题：令 Θ 表示全局 k 维参量空间，例如单变量正态分布即为 $\Theta=(-\infty<\mu<\infty,\ 0<\sigma<\infty)$，令 Θ_1 表示假设 H_1 的参量空间(Θ 的子空间)，例如，若 $X:N(\mu_X,\ \sigma_X^2)$，假设为 $H_1: 3<\mu_X<4$，则 $\Theta_1=(3<\mu_X<4,\ 0<\sigma_X^2<\infty)$。定义检验 $H_1:\boldsymbol{\theta}\notin\Theta_1$ 及备选 H_2：$\boldsymbol{\theta}\notin\Theta_1$ 的检验统计量 Λ 为

$$\Lambda\triangleq\frac{L_{\mathrm{LM}}(\boldsymbol{\theta}^*)}{L_{\mathrm{GM}}(\boldsymbol{\theta}^\dagger)}\tag{7.3-3a}$$

其中 $L_{\mathrm{LM}}(\boldsymbol{\theta}^*)\triangleq\max_{\boldsymbol{\theta}\in\Theta_1}L(\boldsymbol{\theta})$，$L_{\mathrm{GM}}(\boldsymbol{\theta}^\dagger)\triangleq\max_{\boldsymbol{\theta}\in\Theta}L(\boldsymbol{\theta})$。我们可能会问，为什么式(7.3-3a)给出的 Λ 是判决 H_1 是否为真合理的检验统计量。首先，分子的最大化可得在 Θ_1 约束下基于观测的最大似然参量估计，因为搜索是受限于 Θ_1 的，参量子空间的最大化可能不是全局最大，我们称之为局部最大。其次，分母的最大化可得基于观测的无约束的最大似然参量估计，我们称之为全局最大。此处下标LM和GM分别起到提醒读者“局部最大”和“全局最大”作用。注意 Λ 是区间[0, 1]的随机变量(提问读者：为什么?)，若 Λ 趋近于1，则假定 H_1 为真，即未知参量在 Θ_1 中，也可认为是整个空间的最大似然参量。另一方面，若 Λ 很小或趋近于0，则假定最大似然参量不在 Θ_1 中。门限值 c 表示我们接受(拒绝)假设或是接受(拒绝)备选假设 H_2 的分界点，由显著性水平 α 确定。总之，广义似然比检验可以描述为

$$H_1\text{为假},\quad \Lambda<c\tag{7.3-3b}$$

其中 Λ 由式(7.3-3a)给出。可见在一定条件下，广义似然比检验是纽曼-皮尔逊意义下的近似最佳，在文献中也存在广义似然比并不总是最佳的反例[7-22]，从某种意义下它应该视为经验性的。

我们通过几个包括连续分布的例子来阐述广义似然比检验的应用。

例7.3-2　(X 为正态且 σ^2 已知时 $H_1: \mu=\mu_1$ 对 $H_2: \mu\neq\mu_1$ 的检验)已知方差为 σ^2 正态随机变量的 n 个观测，似然函数为

$$\begin{aligned}L(\mu)&=(2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right)\\&=(2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(X_i-\hat{\mu})^2+(\hat{\mu}-\mu)^2\right]\right)\end{aligned}\tag{7.3-4}$$

上式从第一行到第二行我们增加了一些指数交叉项，在合并时会消除，其代数分析留给读者练习。由于 σ^2 给定，空间 Θ 为$(-\infty<\mu<\infty)$，则 $L_{\mathrm{GM}}(\mu^\dagger)$ 由 $\mu^\dagger=\hat{\mu}$ 时得到，即 $L_{\mathrm{GM}}(\mu^\dagger)=L(\hat{\mu})$，由于 Θ_1 仅有满足 $L_{\mathrm{LM}}(\mu^*)=L(\mu_1)$ 的一个点，根据式(7.3-3a)，有

$$\Lambda=\frac{L(\mu_1)}{L(\hat{\mu})}=\exp\left(-\frac{n}{2\sigma^2}(\hat{\mu}-\mu_1)^2\right)\tag{7.3-5}$$

判决域与 $\hat{\mu}$、μ_1 远离程度相关。当 $\hat{\mu}$ 接近 μ_1 时，Λ 将趋近1并判定 H_1 为真，当 $\hat{\mu}$ 远离 μ_1 时，将判定 H_1 为假，介于两者之间 $0<\Lambda<c$ 的常量 c 用来区分判决域。取自然对数后，可得判决域定义为

$$\begin{aligned}\hat{\mu}&>\mu_1+\left(2\sigma^2\ln(1/c)/n\right)^{1/2}\\\hat{\mu}&<\mu_1-\left(2\sigma^2\ln(1/c)/n\right)^{1/2}\end{aligned}\tag{7.3-6}$$

其中 c 由检验的显著性水平 α 确定。

例 7.3-3　(例 7.3-2 的数值实现)根据已知的数值估计式(7.3-6)，假定 $\mu_1=5$，$\sigma^2=4$，$n=15$ 及 $\alpha=0.05$，用 $f_\Lambda(x;\mu_1)$ 表示 Λ 的概率密度函数，我们需要计算 $0.05=\int_0^c f_\Lambda(x)\mathrm{d}x$。由事件 $\{\Lambda\leqslant c\}$ 等效于 $\{-\infty<\ln\Lambda<\ln c\}$，又可等效于 $\{-2\ln c\leqslant -2\ln\Lambda<\infty\}$，根据式(7.3-5)，有 $-2\ln\Lambda=\left(\dfrac{\hat{\mu}-\mu_1}{\sigma/\sqrt{n}}\right)^2$，为 1 自由度的 χ^2 分布，即 χ_1^2(下标表示自由度)。用 $f_{\chi^2}(x;n)$ 表示 χ_n^2 的概率密度函数，可得

$$0.05=\int_0^c f_\Lambda(x)\mathrm{d}x=\int_{-2\log c}^{\infty} f_{\chi^2}(x;1)\mathrm{d}x=1-F_{\chi^2}(-2\ln c;1)$$

从 χ_1^2 分布随机变量的累积密度函数表中可得 $-2\ln c=3.84$。因此根据式(7.3-6)可确定判决域为 $\hat{\mu}>6.01$，$\hat{\mu}<3.99$，或表示为 $\hat{\mu}(-\infty,3.99)\cup(6.01,\infty)$。

例 7.3-4　(电话机排队时等待时间的测试)打电话到投资银行(GIB)，自动(机器)接线员说明在工作时间接入到投资顾问的平均等待时间不超过 30 s。我们将用广义似然比检验来测试。在工作时间打入 n 次电话并记录下等待时间 X_i，$i=1,\cdots,n$，假定等待时间为独立同分布的指数随机变量，其概率密度函数为 $f_{X_i}(x;\mu)=(1/\mu)\exp(-x/\mu)u(x)$，其中 $\mu=E(X_i)$，$i=1,\cdots,n$。从基本概率知识我们知道 $\hat{\mu}=(1/n)\sum_{i=1}^{n}X_i$ 是 μ 的无偏一致估计。检验假设 H_1: $\mu\leqslant0.5$ 对 H_2: $\mu>0.5$，似然函数为 $L(\mu)=(1/\mu)^n\exp\left(-\dfrac{n}{\mu}\left[(1/n)\sum\limits_{i=1}^{n}X_i\right]\right)=(1/\mu^n)\exp(-n\hat{\mu}/\mu)$，则由似然函数对 μ 微分得到 $L_{\mathrm{GM}}(\mu^\dagger)$ 为

$$L_{\mathrm{GM}}(\mu^\dagger)=L(\hat{\mu})=\hat{\mu}^{-n}\exp(-n)$$

计算 $L_{\mathrm{LM}}(\mu^*)$ 会更复杂一些。为更好往下说明，我们将两个似然函数绘制于图 7.3-1 中，一个在均值 0.45 处有峰值，另一个在均值 0.55 处有峰值。μ 的空间 $\Theta_1=(0,0.5]$ 是基于 $\mu\leqslant0.5$ 的假设并包括全局最大点 0.45。然而，当似然函数为图 7.3-1 中下面的曲线时，即峰值位于 $\mu=0.55$ 的曲线，由于 0.55 不包括在 $\Theta_1=(0,0.5]$，局部最大不等于全局最大。

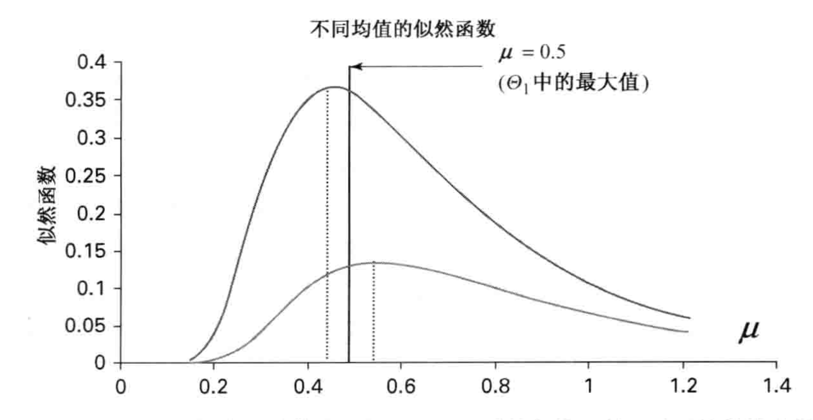


图 7.3-1　上面的曲线为均值为 0.45，$n=10$ 时的似然函数，下面的曲线为均值为 0.55，$n=10$ 时的似然函数。子空间 $\Theta_1=(0,0.5]$ 包括 0.45(显示为实线左边的虚线)，但不包括 0.55(实线右边的虚线)

因此

$$L_{\mathrm{LM}}(\mu^{\dagger})=\begin{cases}\hat{\mu}^{-n}\exp(-n),\hat{\mu}\leqslant 0.5\\ 2^{n}\exp(-2n\hat{\mu}),\hat{\mu}>0.5\end{cases}$$

子空间 $\Theta_1=(0,0.5]$ 包括 0.45(显示为实线左边的虚线),但不包括 0.55(实线右边的虚线)。

最后,根据式(7.3-3a),我们有

$$\Lambda\triangleq\frac{L_{\mathrm{LM}}(\mu^{*})}{L_{\mathrm{GM}}(\mu^{\dagger})}=\begin{cases}1,\ \hat{\mu}\leqslant 0.5\\ (2\hat{\mu})^{n}\exp(-n[2\hat{\mu}-1]),\ \hat{\mu}>0.5\end{cases}\tag{7.3-7}$$

判决域为间隔 $(0,c')$,即 $\Lambda\in(0,c')$ 将判决 H_1 为假。判决域由图 7.3-2 绘出,从 Λ 轴看其小于水平线 c',从 $\hat{\mu}$ 轴看为 $\hat{\mu}=c$ 的右边。

因为似然函数在区域 $\hat{\mu}>0.5$ 随 $\hat{\mu}$ 单调递减(参见图 7.3-2),可以采用 $\hat{\mu}$ 为检验统计量。假定 n 足够大,以至于 $\hat{\mu}$ 近似正态,即概率密度函数是以均值为中心且方差很小,由于 $\hat{\mu}$ 是 μ 的无偏估计,有 $\hat{\mu}$: $N(\mu,\mu^2/n)$,这个结果应用到单一指数随机变量的方差为 μ^2,因此 $\hat{\mu}$ 的方差为 $\sigma_{\hat{\mu}}^2=\mu^2/n$。我们由 $Z\triangleq(\hat{\mu}-\mu)\sqrt{n}/\mu$ 得到近似标准正态随机变量,c 由显著性约束 α 计算得出。采用百分数表示 $1-\alpha=F_{\mathrm{SN}}(z_{1-\alpha})$,可得 $c=\mu+z_{1-\alpha}\mu/\sqrt{n}$,可见判决点 c 随 μ 线性增加。当 $\hat{\mu}>c$ 时拒绝假设,例如当 $\mu=0.5$,$\alpha=0.05$ 及 $n=10$ 时可得 $z_{0.95}=1.64$ 及 $c\approx0.76$,当 α 增加,分割点向 0.5 处下降(参见图 7.3-3)。

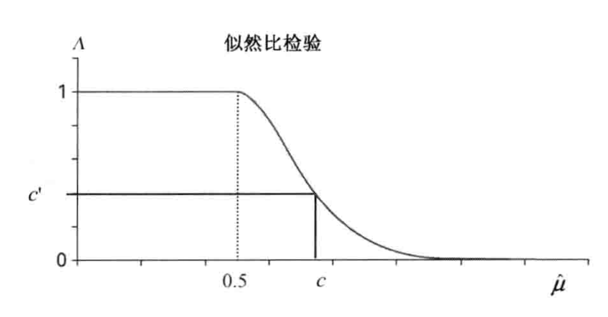


图 7.3-2 广义似然比检验统计量与样本均值估计的变化

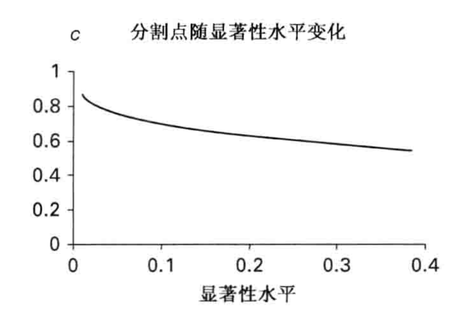


图 7.3-3 当 α 增加,分割点随之下降,增加了判决域的宽度

例 7.3-5 (评估药物赫塞汀的肿瘤治疗效果)对肿瘤治疗的新疗法包括使对肿瘤细胞提供能量的蛋白质无效。例如,一些乳腺癌细胞含有蛋白质 HER2,药物赫塞汀可以降低癌细胞 50% 活跃度,达到治疗效果①,而没有 HER2 的肿瘤其后果要乐观得多。赫塞汀有很大的毒副作用,因此对蛋白质 HER2 的检验必须很精确,尽管常常没有做到。令 H_1:肿瘤有高含量的 HER2,对赫塞汀有反应。令 H_2:肿瘤有低含量(或没有)的 HER2,病人不需要赫塞汀。当前对 HER2 的测试为

$$P[H_1\text{为真}|H_2\text{为真}]=0.2$$
$$P[H_2\text{为真}|H_1\text{为真}]=0.1$$

因而,检验的显著性水平为 0.1,检测的势为 0.8。

① “癌症战争:新药的不确定性测试”,纽约时报,8 月 20 日,2010 年。

如何检验两个集合的均值相等

假定对一类治疗前列腺癌药物进行测试①，测试者为重度无雄性荷尔蒙的前列腺癌患者，与常规治疗相比，这种药物能不能延长参与测试者的生命呢？一个印刷公司评估其印刷机应用的两种纸，是否一种会比另一种纸更容易使印刷机卡住呢？运输部门考虑购买两类水泥，是否其中一类更不容易产生凹坑呢？这类问题都符合下述架构，即我们有两类正态分布集合，从集合 P_1 中取 m 个样本，从集合 P_2 中取 n 个样本，集合 P_1 与集合 P_2 的均值是否相同？通常，这是一个困难的问题，特别是这超越了本章探讨的范围。更多关于此问题的讨论参见参考文献[7-1]。然而，当假定集合的方差相同时，这个问题可以直接解析得到。为对此问题的讨论，先通过例 7.3-6 复习一下相关内容。

例 7.3-6 （例 7.3-7 的初步结论）两个正态集合为 $S_1=\{X_{1i},\ i=1,\ \cdots,\ m\}$ 和 $S_2=\{X_{2i},\ i=1,\ \cdots,n\}$，$S_1$ 中元素为 m 个满足 $X_1:N(\mu_1,\ \sigma_1^2)$ 的 X_1 的独立同分布观测，同样，S_2 中元素为 n 个满足 $X_2:\ N(\mu_2,\ \sigma_2^2)$ 的 X_2 的独立同分布观测。另外，假定对于任意 i,j，满足 $E[(X_{1i}-\mu_1)\cdot(X_{2j}-\mu_2)]=0$。

(i) 假定 $\mu_1=\mu_2=\mu$，证明 $E[\hat{\mu}_1-\hat{\mu}_1]=0$。

(i)的解答： $E[\hat{\mu}_1-\hat{\mu}_2]=E[\hat{\mu}_1]-E[\hat{\mu}_2]=\mu-\mu=0$

(ii) 假定 $\mu_1=\mu_2=\mu$ 且 $\sigma_1=\sigma_2=\sigma$，证明 $\text{Var}=(\hat{\mu}_1-\hat{\mu}_2)=(m^{-1}+n^{-1})\sigma^2$。

(ii)的解答： 由于 $E[\hat{\mu}_1]=E[\hat{\mu}_2]=\mu$，$\text{Var}(\hat{\mu}_1-\hat{\mu}_2)=E[(\hat{\mu}_1-\hat{\mu}_2)^2]=E[\hat{\mu}_1^2]+E[\hat{\mu}_2^2]-2E[\hat{\mu}_1\hat{\mu}_2]$。将

$$
\begin{aligned}
&E[\hat{\mu}_1^2]=m^{-2}\left(\sum_{i=1}^{m}E(X_{1i}^2)+\sum_{i=1}^{m}\sum_{j\neq i}^{m}E(X_{1i}X_{1j})\right)\\
&E[\hat{\mu}_2^2]=n^{-2}\left(\sum_{i=1}^{n}E(X_{2i}^2)+\sum_{i=1}^{n}\sum_{j\neq i}^{n}E(X_{2i}X_{2j})\right)\\
&E(X_{1i}^2)=\mu^2+\sigma^2=E(X_{2i}^2)\\
&E(\hat{\mu}_1\hat{\mu}_2)=\mu^2
\end{aligned}
$$

代入方差表达式可得所需结论。

(iii) 若 V 和 W 分别为自由度为 m 和 n 的卡方分布，则 $U\triangleq V+W$ 为自由度为 $m+n$ 的卡方分布。

(iii)的解答： 若 $V:\chi_m^2$ 及 $W:\chi_n^2$，则有 $V=\sum_{i=1}^{m}Y_i^2$ 和 $W=\sum_{i=1}^{n}Z_i^2$，其中我们可以假定 $Y_i,\ i=1,\ \cdots,\ m$ 为满足 $N(0,\ 1)$ 的独立同分布变量，$Z_i,\ i=1,\ \cdots,\ n$ 为满足 $N(0,\ 1)$ 的独立同分布变量，V 的 MGF 为 $M_V(t)=E[\exp(tV)]$，由下式计算

$$
\begin{aligned}
M_V(t)&=(2\pi)^{-m/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left(t\sum\nolimits_{i=1}^{m}y_i^2\right)\times\exp\left(-1/2\sum\nolimits_{i=1}^{m}y_i^2\right)\prod\nolimits_{i=1}^{m}\mathrm{d}y_i\\
&=\prod\nolimits_{i=1}^{m}(2\pi)^{-1/2}\int_{-\infty}^{\infty}\exp\left(-0.5(1-2t)y_i^2\right)\mathrm{d}y_i\\
&=(1-2t)^{-m/2},\quad t<1/2
\end{aligned}
$$

第一行是定义，第二行应用了 Y_i 的独立同分布假定，第三行是依据正态曲线下积分面积为 1

---

① 无雄性荷尔蒙意味着癌细胞不是由睾丸素提供能量，这很难治疗。作者的同事，Nick Galatsanos 教授，对图像处理做出重要贡献，于 52 岁时死于此疾病。

得到的。由于 $U=V+W$ 及 V 和 W 是相互独立的，从而由4.4节内容有 $M_U(t)=M_V(t)m_V(t)$。由 $M_V(t)=(1-2t)^{-m/2}$ 及 $M_W(t)=(1-2t)^{-n/2}$，可得 $M_U(t)=(1-2t)^{-(m+n)/2}$，即说明 U: χ^2_{m+n}。

(iv)给定似然函数 $L=(2\pi\sigma^2)^{-m/2}\exp[-0.5\sum_{i=1}^{m}((X_i-\mu)^2/\sigma^2)]$，证明 $L_{GM}=L(\hat{\mu}^{\dagger}$[1]，$\hat{\sigma}^{2\dagger})$，其中 $\hat{\mu}^{\dagger}=\hat{\mu}$ 及 $\hat{\sigma}^{2\dagger}=\hat{\sigma}^2$。

(iv)的解答： $\hat{\mu}^{\dagger}$ 可通过将 $\ln L$ 对 μ 微分得到，为 $\hat{\mu}^{\dagger}=\hat{\mu}\triangleq(m)^{-1}\sum_{i=1}^{m}X_i$，同样，$\hat{\sigma}^{2\dagger}$ 可通过将 $\ln L$ 对 σ^2 微分得到，为 $\hat{\sigma}^{2\dagger}=\hat{\sigma}^2\triangleq m^{-1}\sum_{i=1}^{m}(X_i-\hat{\mu})^2$，代入 L 的表达式计算 L_{GM} 得到

$$L_{GM}=\left(\frac{m}{2\pi\sum_{i=1}^{m}(X_i-\hat{\mu})^2}\right)^{m/2}e^{-m/2} \tag{7.3-8}$$

例7.3-7 （$H_1:\mu_1=\mu_2$ 对 $H_2:\mu_1\neq\mu_2$，$\sigma_1^2=\sigma_2^2=\sigma^2$ 未知时的检验）如同例7.3-6中，我们有两个正态样本集合 $S_1=\{X_{1i},i=1,\cdots,m\}$ 和 $S_2=\{X_{2i},i=1,\cdots,n\}$，$S_1$ 中元素为 m 个满足 X_1: $N(\mu_1,\sigma_1^2)$ 的 X_1 的独立同分布观测，同样，S_2 中元素为 n 个满足 X_2: $N(\mu_2,\sigma_2^2)$ 的 X_2 的独立同分布观测。另外，假定对任意 i 和 j，有 $E[(X_{1i}-\mu_1)(X_{2j}-\mu_2)]=0$。我们要检验 $H_1:\mu_1=\mu_2$ 对 $H_2:\mu_1\neq\mu_2$，H_1 的参数空间[1]为 $\Theta_1=(\mu,\sigma^2)$，全局参数 空间为 $\Theta=(\mu_1,\mu_2,\sigma^2)$，似然函数为

$$L=\left(\frac{1}{2\pi\sigma^2}\right)^{(m+n)/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left(\frac{X_{1i}-\mu_1}{\sigma}\right)^2\right)\times\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_{2i}-\mu_2}{\sigma}\right)^2\right) \tag{7.3-9}$$

根据例7.3-6中(iv)的结论，可得

$$\hat{\mu}_1^{\dagger}=m^{-1}\sum_{i=1}^{m}X_{1i}=\hat{\mu}_1$$
$$\hat{\mu}_2^{\dagger}=n^{-1}\sum_{i=1}^{n}X_{2i}=\hat{\mu}_2$$
$$\hat{\sigma}^{2\dagger}=\tfrac{1}{m+n}\left(\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2\right)=\hat{\sigma}^2$$

我们将此结果分别代入 L 中的 μ_1，μ_2 及 σ^2 中，可得

$$L_{GM}=\left(\frac{m+n}{2\pi\left(\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2\right)}\right)^{(m+n)/2}\exp\left(-\frac{(m+n)}{2}\right) \tag{7.3-10}$$

回到式(7.3-9)的似然函数，我们希望在参数空间 Θ_1 最大化此式，由于 H_1 中有 $\mu_1=\mu_2=\mu$，L 写为

$$L(\mu,\sigma)=\left(\frac{1}{2\pi\sigma^2}\right)^{(m+n)/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left(\frac{X_{1i}-\mu}{\sigma}\right)^2\right)\times\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_{2i}-\mu}{\sigma}\right)^2\right) \tag{7.3-11}$$

分别对 μ 和 σ^2 微分得到 $\hat{\mu}^*$ 和 $\hat{\sigma}^{2*}$ 为

$$\begin{aligned}\hat{\mu}^*&=\frac{1}{m+n}\left(\sum_{i=1}^{m}X_{1i}+\sum_{i=1}^{n}X_{2i}\right)\\&=\frac{m}{m+n}\hat{\mu}_1+\frac{n}{m+n}\hat{\mu}_2\end{aligned}$$

---

[1] 为避免使用过多的记号且不会引起混淆时，我们用 $\Theta=\{\mu_1,\mu_2,\sigma_1,\sigma_2\}$ 等方式来表示参量空间 $\Theta=\{-\infty<\mu_1<\infty,-\infty<\mu_2<\infty,\sigma_1>0,\sigma_2>0\}$，表达式 $L(\Theta)$ 可以解释为在参数空间 Θ 的似然函数。

和

$$\hat{\sigma}^{2*}=\frac{1}{m+n}\left(\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2+\frac{mn}{m+n}(\hat{\mu}_1-\hat{\mu}_2)^2\right)$$

将$\hat{\mu}^*$和$\hat{\sigma}^{2*}$代入式(7.3-11)的$L(\Theta_1)$中的μ和σ^2，可得L_{LM}为

$$L_{LM}=\left[\frac{(m+n)\mathrm{e}^{-1}}{2\pi\left(\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2+\frac{mn}{m+n}(\hat{\mu}_1-\hat{\mu}_2)^2\right)}\right]^{(m+n)/2}$$

似然比$\Lambda\triangleq L_{LM}/L_{GM}$为

$$\Lambda=\left[1+\frac{\frac{mn}{m+n}(\hat{\mu}_1-\hat{\mu}_2)^2}{\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2}\right]^{-(m+n)/2} \tag{7.3-12}$$

从例7.3-6的(ii)的结论，$\hat{\mu}_1-\hat{\mu}_2$的分布满足$N(0,\sigma^2(m+n)/mn)$，因此

$$Z\triangleq\frac{(\hat{\mu}_1-\hat{\mu}_2)}{\sigma\left(\frac{m+n}{mn}\right)^{1/2}}$$

满足$N(0,1)$分布。同样地，根据例7.3-6中(iii)的结论，

$$W_{m+n-2}\triangleq\left(\sum_{i=1}^{m}\left(\frac{X_{1i}-\hat{\mu}_1}{\sigma}\right)^2+\sum_{i=1}^{n}\left(\frac{X_{2i}-\hat{\mu}_2}{\sigma}\right)^2\right)$$

为自由度$m+n-2$的卡方分布。最后，回想$T_{m+n-2}=\dfrac{Z\sqrt{m+n-2}}{W_{m+n-2}}$为自由度为$m+n-2$的$t$分布随机变量，因此

$$\Lambda=\left[1+\left(T_{m+n-2}^2/(m+n-2)\right)\right]^{-(m+n)/2} \tag{7.3-13}$$

由于Λ是T_{m+n-2}^2的单调递减函数，根据T_{m+n-2}^2检验比Λ检验要好，则H_1的判决域$0<\Lambda<\lambda_c$，当由T_{m+n-2}^2检验时变为(t_c^2,∞)(参见图7.3-4)，或等效为事件(t_c,∞)和$(-\infty,-t_c)$的并集(参见图7.3-5)。更多关于称为t检测的此类检验可参见参考文献[7-10]至参考文献[7-15]，也可以在Google或其他搜索引擎上输入"t检测"获得。

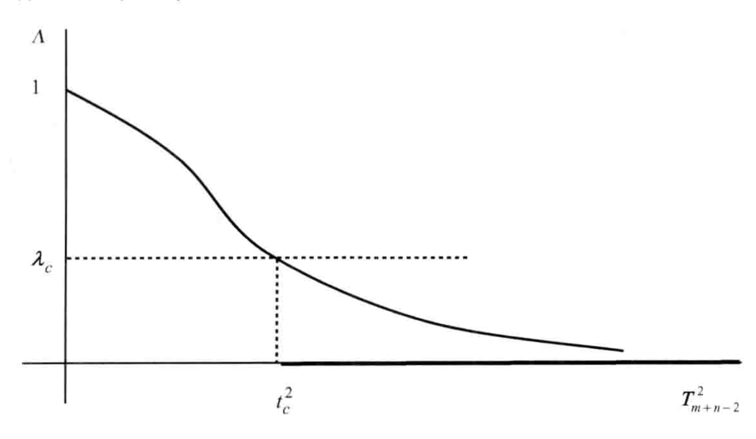


图7.3-4 判决域用粗线条显示，应用T_{m+n-2}^2比Λ更容易检验H_1对H_2

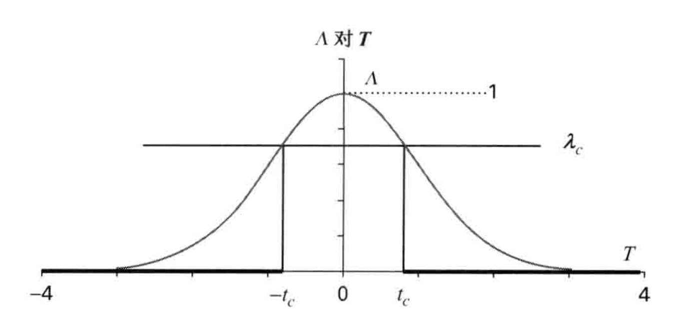


图7.3-5 相比较于广义似然比统计量，T统计量更适合检验，参见式(7.3-13)。沿T轴的判决域显示为粗线条。依据读者的兴趣，此图形中$m=n=10$。若$|T|>t_c$，假设为假，其中t_c依赖于第一类错误α。在显著性α的双边检验中，我们为每边判决域指定α/2错误概率，即$P[T>t_c]=\alpha/2$和$P[T<-t_c]=\alpha/2$

在第一类错误α约束下，若有$\{T^2>t^2_{1-\alpha/2}\}$，则判决假设不成立，其中$t_{1-\alpha/2}$是从t分布表中选取自由度$m+n-2$并根据$F_T(t_{1-\alpha/2})=1-\alpha/2$获得。

例7.3-8 （检验$H_1:\mu_1=\mu_2$对$H_2:\mu_1\neq\mu_2$的数值实例）应用高斯随机数发生器（可从互联网获取）生成15个满足$N(0,2)$的样本集合(P_1)，15个满足$N(2,2)$的样本集合(P_2)，得到的数值如下：

集合P_1：$S_1=\{2.21, 0.83, 0.393, 0.975, 0.195, -0.069, -1.91, 1.44, -3.98, 0.98, -2.84, -1.56, -0.4, -1.08, 0.116\}$；$\hat{\mu}'_1=-0.258$；$m=15$；$\sum_{i=1}^{15}(X'_{1i}-\hat{\mu}'_1)^2=40.88$。

集合P_2：$S_2=\{-1.28, -0.258, -0.947, 5.85, 1.56, 1.48, 1.95, 3.22, 1.41, 1.84, 2.69, 3.94, 2.04, 2.08, 1.44\}$；$\mu'_2=1.801$；$n=15$；$\sum_{i=1}^{15}(X'_{2i}-\hat{\mu}'_2)^2=45.66$。将数值代入

$$T^2\triangleq(m+n-2)\frac{mn(m+n)^{-1}(\hat{\mu}_1-\hat{\mu}_2)^2}{\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2} \tag{7.3-14}$$

得到T^2的值为10.34。最后，根据$\alpha=P(H_1\text{为假}|H_1\text{为真})=0.01$，可得$F_T(t_{1-\alpha/2})=1-0.005=0.995$，自由度为$15+15-2=28$。根据t分布表可得$t_{1-\alpha/2}=2.763$，即$t^2_{1-\alpha/2}=7.63$，由于$T^2>t^2_{1-\alpha/2}$，我们判定均值相等的假设不成立。

正态集合方差相等的检验：F检验

我们遇到的另一个问题是：两个正态集合是否有相同的方差。模型如下：有两个正态集合P_1，$N(\mu_1,\sigma_1^2)$和P_2，$N(\mu_2,\sigma_2^2)$，并从P_1中得到m个样本$S_1=\{X_{1i}, i=1,\cdots,m\}$（即m个独立同分布观测），从$P_2$中得到n个样本$S_2=\{X_{2i}, i=1,\cdots,n\}$。基于样本我们希望检验假设$H_1:\sigma_1^2=\sigma_2^2\triangleq\sigma^2$对备选$H_2:\sigma_1^2\neq\sigma_2^2$，$H_1$检验的参数空间为$\Theta_1=\{\mu_1,\mu_2,\sigma^2\}$，$H_2$检验的参数空间为$\Theta=\{\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\}$，似然函数为

$$L(\Theta)=(2\pi\sigma_1^2)^{-m/2}\exp\left(-0.5\sum_{i=1}^{m}\left(\frac{X_{1i}-\mu_1}{\sigma_1}\right)^2\right)$$
$$\times(2\pi\sigma_2^2)^{-n/2}\exp\left(-0.5\sum_{i=1}^{n}\left(\frac{X_{2i}-\mu_2}{\sigma_2}\right)^2\right)$$

在$\Theta_1=\{\mu_1,\mu_2,\sigma^2\}$中形式为

$$L(\Theta_1)=(2\pi\sigma^2)^{-(m+n/2)}\exp\left(-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{m}(X_{1i}-\mu_1)^2+\sum_{i=1}^{n}(X_{2i}-\mu_2)^2\right]\right)$$

在 Θ_1 使得 $L(\Theta)$ 最大的参数，和前述一样，通过 $\ln L(\Theta_1)$ 对 μ_1，μ_2，σ^2 的微分并令其等于零得到

$$\hat{\mu}_1^*=(m)^{-1}\sum_{i=1}^{m}X_{1i}=\hat{\mu}_1;\hat{\mu}_2^*=(n)^{-1}\sum_{i=1}^{n}X_{2i}=\hat{\mu}_2$$

$$\sigma^{2*}=(m+n)^{-1}\left(\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2\right)$$

将此结果代入 $L(\Theta_1)$ 可得

$$L_{\mathrm{LM}}=\left(\frac{2\pi}{(m+n)}\left[\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2\right]\right)^{-(m+n)/2}\exp\left(-(m+n)/2\right)$$

在 $\Theta=\{\mu_1,\mu_2,\sigma_1^2,\sigma_2^2\}$ 对 $L(\Theta)$ 最大化，通过 $\log L(\Theta)$ 对 μ_1，μ_2，σ_1^2，σ_2^2 的微分并令其等于零得到

$$\hat{\mu}_1^\dagger=(m)^{-1}\sum_{i=1}^{m}X_{1i}=\hat{\mu}_1;\hat{\mu}_2^\dagger=(n)^{-1}\sum_{i=1}^{n}X_{2i}=\hat{\mu}_2$$

$$\hat{\sigma}_1^{2\dagger}=(m)^{-1}\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2=\hat{\sigma}_{1,ML}^2;\hat{\sigma}_2^{2\dagger}=(n)^{-1}\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2=\hat{\sigma}_{2,ML}^2$$

注意到方差 σ_1^2，σ_2^2 的最大似然估计 $\hat{\sigma}_{1,\mathrm{ML}}^2$，$\hat{\sigma}_{2,\mathrm{ML}}^2$是无偏估计，将结果代入 $L(\Theta)$，得到L_{GM}为

$$L_{\mathrm{GM}}=(2\pi)^{-(m+n)/2}\left(\frac{1}{m}\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2\right)^{-m/2}\left(\frac{1}{n}\sum_{i=1}^{m}(X_{2i}-\hat{\mu}_2)^2\right)^{-n/2}$$
$$\times\exp\left(-(m+n)/2\right)$$

最后，由 $\Lambda=L_{\mathrm{LM}}/L_{\mathrm{GM}}$可得

$$\Lambda=\frac{\left(\dfrac{m+n}{\left(\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2+\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2\right)}\right)^{(m+n)/2}}{\left(\dfrac{m}{\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2}\right)^{m/2}\left(\dfrac{n}{\sum_{i=1}^{n}(X_{2i}-\hat{\mu}_2)^2}\right)^{n/2}}\tag{7.3-15}$$

这个复杂的表达式可以通过下述结论进行简化

$$(m-1)\hat{\sigma}_1^2=\sum_{i=1}^{m}(X_{1i}-\hat{\mu}_1)^2$$
$$(n-1)\hat{\sigma}_2^2=\sum_{j=1}^{n}(X_{2j}-\hat{\mu}_2)^2$$

经过代数变换，可得

$$\Lambda=A(m,n)\frac{\left([(m-1)/(n-1)]\times\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}\right)^{m/2}}{\left(1+[(m-1)/(n-1)]\times\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}\right)^{(m+n)/2}}$$

其中 $A(m,n)\triangleq(m+n)^{(m+n)/2}m^{-m/2}n^{-n/2}$。我们将 $V_R\triangleq\hat{\sigma}_1^2/\hat{\sigma}_2^2$ 自然地称为(估计)方差比，其中

$$V_R\triangleq\frac{(m-1)^{-1}\sum_{i=1}^{m}(X_i-\hat{\mu}_1)^2}{(n-1)^{-1}\sum_{i=1}^{n}(X_i-\hat{\mu}_2)^2}\tag{7.3-16}$$

则用 V_R 表示有

$$\Lambda=A(m,n)\frac{([(m-1)/(n-1)]\times V_R)^{m/2}}{(1+[(m-1)/(n-1)]\times V_R)^{(m+n)/2}}\triangleq\Lambda(V_R)\tag{7.3-17}$$

当H_1为真有$V_R = F_{m-1, n-1}$，其中$F_{m-1, n-1}$为自由度分别为$m-1$和$n-1$的F分布随机变量。当$m=n=10$时Λ与V_R的变化显示在图7.3-6中，从图可见，判决假设不成立的事件$\{0<\Lambda(V_R)<c\}$等价于两个拖尾事件$\{0<V_R<t_l\} \cup \{t_u<V_R<\infty\}$。当给定显著性水平$\alpha$，根据$P[0<V_R<t_l]+P[t_u<V_R<\infty]=\alpha$及$\Lambda(t_l)=\Lambda(t_u)$可得到$t_l'$和$t_u'$。在不损失太多精度情况下简化分析，选择$P[0<V_R<t_l']=P[t_u'<V_R<\infty]=\alpha/2$，$t_l'$和$t_u'$的数值比$t_l$和$t_u$容易确定，参见图7.3-7。实际上，用$F_F(x_\beta; m-1; n-1)$来表示随机变量$F_{m-1, n-1}$在$\beta$百分点处的累积密度函数，即$F_F(x_\beta; m-1; n-1) \triangleq \beta$，可以得到$t_l'=x_{\alpha/2}$和$t_u'=x_{1-\alpha/2}$。

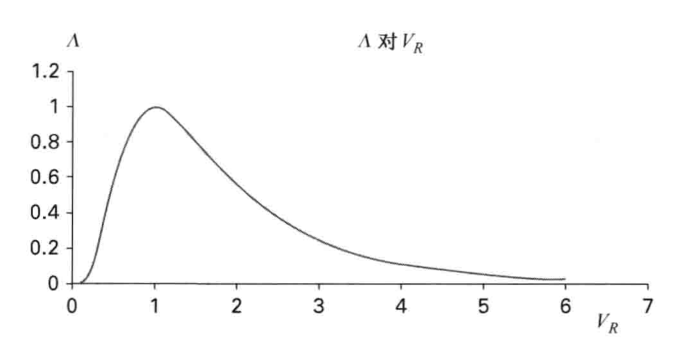


图7.3-6 Λ时检验统计量V_R与方差比$m=n=10$

当检验为$\{0<\Lambda<c\}$，或等效于$\{0<V_R<x_{\alpha/2}\}$或$\{x_{1-\alpha/2}<V_R\}$时，假设H_1不成立。

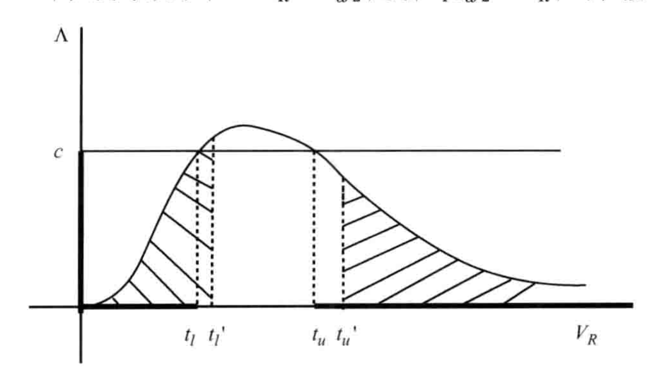


图7.3-7 事件$\{0<\Lambda<c\}$等效于事件$\{0<V_R<t_l\} \cup \{t_u<V_R<\infty\}$，数值$t_l$和$t_u$由$t_l'$和$t_u'$取代，使得两边的错误均为$\alpha/2$

例7.3-9 (检验H_1: $\sigma_1^2=\sigma_2^2 \triangleq \sigma_0^2$对$H_2$: $\sigma_1^2 \neq \sigma_2^2$的数值实例)检验两个集合的方差相同的假设。

应用互联网上的RANDOM. ORG程序生成两组高斯伪随机数如下：

$N(0,1)$:	0.436,	−1.06,	−1.11,	0.46,	0.491,	−1.05,	0.502,	0.598,	1.61
				−0.981,	−0.021,	0.253,	−1.24,	0.059,	2.12
$N(0,4)$:	0.634,	0.0818,	−1.32,	2.96,	3.11,	3.13,	2.62,	−1.96,	0.85
				−6.51,	−3.39,	4.25,	−1.08,	3.42,	2.72

从前两行，即($N(0, 1)$)数据，计算得到$\hat{\mu}_1=0.074$，$\hat{\sigma}_1=1.01$，$\hat{\sigma}_1^2=1.01$；从下面两行，即($N(0, 4)$)数据，可计算得到$\hat{\mu}_2=0.54$，$\hat{\sigma}_2=3.04$，$\hat{\sigma}_2^2=9.25$，计算得到的方差比为

$$V_R = \frac{(15-1)\sum_{i=1}^{15}(X_{1i}-0.54)^2}{(15-1)\sum_{i=1}^{15}(X_{2i}-0.074)^2} = \frac{9.25}{1.02} = 9.06$$

约束显著性水平 $\alpha=0.05$，应用“等区域”系统来分配错误概率，我们计算满足 $F_F(x_{0.025};14;14)=0.025$ 和 $F_F(x_{0.975};14;14)=0.975$ 的百分点 $x_{0.025}$ 及 $x_{0.975}$。等效于应用 F 分布表，我们调用 Stat Trek 在线统计表中 F 分布计算器，输入自由度(两种情况均为 14)及累积密度函数值 0.025，得到 $x_{0.025}=0.34$，重复并输入累积密度函数值 0.975 得到 $x_{0.975}=2.98$，因此，接受区域为间隔(事件)(0.34, 2.98)，拒绝区域为事件$\{(0,0.34)\cup(2.98,\infty)\}$。检验统计量为 9.06 落在拒绝区域，因此两个方差相同的假设不成立，我们可以相当肯定数据来自两个不同集合的结论。

更多称为 F 检验内容可以从参考文献[7-5]至参考文献[7-9]获得，也可用 Google 在线输入“F 检验”搜索得到。

正态集合是否预设方差值的检验

在此我们考虑一个正态集合并检验其是否预设了方差值。过程如下：从正态集合 X 中取 m 个样本，即从 X 中取得 m 个独立同分布观测并记为$\{X_i,\ i=1,\cdots,m\}$，H_1假定集合的方差预设为 σ_0^2，备选假设 H_2假定集合方差不等于 σ_0^2，或更精确地，没有足够证据表明 H_1成立。与前面一样，分析似然函数并令其在 $\Theta_1=\{\mu,\sigma_0^2\}$ 和 $\Theta=\{\mu,\sigma^2\}$ 最大化，因而 $L(\Theta_1)=(2\pi\sigma_0^2)^{-m/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left(\frac{X_i-\mu}{\sigma_0}\right)^2\right)$，即当 $\hat{\mu}^*=\hat{\mu}\triangleq(m)^{-1}\sum_{i=1}^{m}X_i$ 时的最大值。有

$$L_{\mathrm{LM}}=(2\pi\sigma_0^2)^{-m/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left(\frac{X_i-\hat{\mu}}{\sigma_0}\right)^2\right)$$

同样

$$L(\Theta)=(2\pi\sigma^2)^{-m/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left(\frac{X_i-\mu}{\sigma}\right)^2\right)$$

是当 $\hat{\mu}^{\dagger}=\hat{\mu}\triangleq(m)^{-1}\sum_{i=1}^{m}X_i$ 和 $\hat{\sigma}^{2\dagger}=\hat{\sigma}^2=(m)^{-1}\sum_{i=1}^{m}(X_i-\hat{\mu})^2$ 时的最大值，因而

$$L_{\mathrm{GM}}=(2\pi\hat{\sigma}^2)^{-m/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left(\frac{X_i-\hat{\mu}}{\hat{\sigma}}\right)^2\right)$$

广义似然比为

$$\begin{aligned}\Lambda&=L_{\mathrm{LM}}/L_{\mathrm{GM}}\\&=\left((m)^{-1}\sum_{i=1}^{m}\left(\frac{X_i-\hat{\mu}}{\sigma_0^2}\right)^2\right)^{m/2}\exp\left(-0.5\sum_{i=1}^{m}\left(\frac{X_i-\hat{\mu}}{\sigma_0^2}\right)^2+m/2\right)\end{aligned}$$

注意到 $W\triangleq\sum_{i=1}^{m}((X_i-\hat{\mu})/\sigma_0)^2$ 服从 χ_{m-1}^2 分布。因此

$$\Lambda=\left((m)^{-1}W\right)^{m/2}\exp\left(-0.5\left(W-m\right)\right)$$

当 DOF =9 时 W 对 Λ 的图形绘于图 7.3-8 中。

从图 7.3-8 中我们可以推断判决事件$\{0<\Lambda<c\}$等效于事件$\{0<W<t_l\}\cup\{t_u<W<\infty\}$，其中 $\Lambda(t_l)=\Lambda(t_u)$ 及 $t_l<t_u$。为了简化，可以选择“等区域”准则，即 $t_l'<t_u'$数值满足 $t_l'=x_{\alpha/2}$及 $t_u'=x_{1-\alpha/2}$，其中 $x_{\alpha/2}$ 和 $x_{1-\alpha/2}$ 为 $\alpha/2$ 和 $1-(\alpha/2)$ 百分点，满足 $F_{\chi^2}(x_{\alpha/2};m-1)=\alpha/2$ 和 $F_{\chi^2}(x_{1-\alpha/2};m-1)=1-(\alpha/2)$，同前，$\alpha=P[H_1\text{为假}|H_1\text{为真}]$。

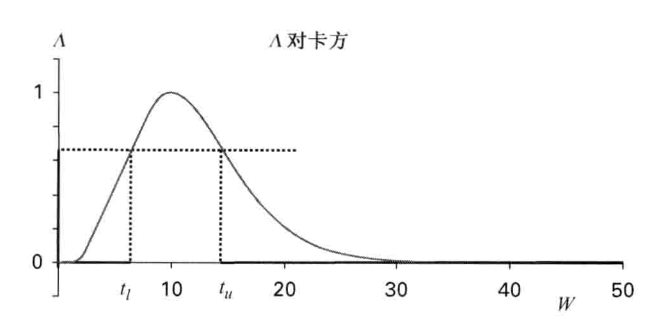


图 7.3-8 Λ 的拒绝区域显示为沿纵坐标的粗线，与双边 W 的拒绝区域相关(显示为沿横坐标的粗线)

例 7.3-10 (检验 $H_1: \sigma^2=\sigma_0^2$ 对 $H_2: \sigma^2\neq\sigma_0^2$ 的数值实例)为达到检验目的，分别从 P_1 和 P_2 两个正态随机数中抽取两组数据，P_1 集合为 $N(1, 1)$，P_2 集合为 $N(1, 4)$。我们检验两个集合 $\sigma^2=1$ 的假设，数值由互联网上 RANDOM. ORG 程序生成

$N(1,1)$ $[P_1]$	−0.0644	2.91	−0.323	1.21	2.66	0.45	1.26	0.923	1.96	1.62
$N(1,4)$ $[P_2]$	0.705	0.685	0.718	1.03	2.52	1.96	0.417	2.69	−1.52	2.98

从 P_1 数据可得 $W'=10.3$，在显著性水平 0.05 时的拒绝区域为 $\{0<W<2.7\}\cup\{19<W<\infty\}$，由于 W 在拒绝区域外，判定 P_1 方差为 1 的假设为真。应用 P_2 重复此过程，计算得到 $W'=16.5$，依然在拒绝区域外(勉强)，我们依然(错误的)判定 P_1 方差为 1 的假设为真。在显著性水平 0.2 重复此过程，拒绝区域为 $\{0<Z<4.7\}\cup\{14.7<Z<\infty\}$，发现判定 P_1 方差为 1 的假设依然为真，但判定 P_2 方差为 1 的假设为假。从此例中可得两点结论：(1)小数据量可能导致错误，其结论应以怀疑的眼光来看；(2)考虑到 α 的含义，我们发现若此参数选得非常小，判决域变小，假设为假的可能性非常小。

7.4 拟合度

统计中很重要的一个问题是检验一个概率集合是否有预设的特定值，例如，假定我们希望确定观测数据是否来自标准正态分布，则函数正态表 F_{SN} 计算概率 $p_i=F_{SN}(x_{i+1})-F_{SN}(x_i)$，$i=1, \cdots, l$，将此数值与随机变量 X 的多个独立观测的数据比较。用此方法可以检验其他分布，不管是离散还是连续的。通用模型是将数据排序进 l 个“容器”，并将估计概率 $\hat{p}_i(i=1, \cdots, l)$ 与特定概率 p_i 进行比较。若 Y_i 表示 n 次试验结果并进入第 i 个“容器”的数据，则有 $\hat{p}_i=Y_i/n$。若所有 $\{\hat{p}_i\}$ 与对应的 $\{p_i\}$ 十分相近，则集合的数据很可能有预定概率。然而，若两个或更多个 $\hat{p}_i$ 与对应 p_i 相差较大，则不能得到测试集合与假定的有相同参数的结论。“容器”数目 l 是一离散随机变量且数目有限(即样本空间的单元数目)，其选择通常就是结果的数目，即对掷骰子而言，l 选择6，对掷硬币，l 为2。对于连续随机变量，“容器”转化为间隔(x_i, x_{i+1}) $i=1, \cdots, l$，即形式为 $\{x_i<X<x_{i+1}, i=1, \cdots, l\}$ 的 l 项结果，l 的选择需要更多考虑。我们需要一个多“精炼的”检验呢？一个精炼的检验将有很多容器，通常数据比容器数目大得多，获取如此多的数据可能花费较多且不切实际，然而，若我们选择做“不精炼的”检验，即选择很小的容器数，则检验必然是粗糙的。另一方面，一个非常大的容器数目而没有足够数据将导致错误百出，检验变得毫无意义。

这些考虑更准确地说，是实验设计和数据处理需要考虑的，而不在本书研究的范畴内。

在拟合度检验中，假设H_1为概率集合$\{p_i,\ i=1,\ \cdots,\ l\}$满足$\{p_i=p_{0i},\ i=1,\ \cdots,\ l\}$，给定概率$\{p_{0i},\ i=1,\ \cdots,\ l\}$表示了概率函数的特征，如分布函数、掷骰子的输出概率，等等，我们获取随机变量X的n个独立同分布观测并根据其数值排序进l个容器中。

定义随机变量X_{ij}为

$$X_{ij} \triangleq \begin{cases} 1, & \text{当}X\text{的第}j\text{次观察在容器}i\text{时} \\ 0, & \text{其他} \end{cases}$$

定义$P[X_{ij}=1]\triangleq p_i$，$i=1,\ \cdots,\ l$，其因独立同分布约束而独立于$j$。

随机变量

$$Y_i \triangleq \sum_{j=1}^{n} X_{ij}, i=1,\cdots,l$$

表示n次试验中在容器$i=1,\ \cdots,\ l$的值。注意有$\sum_{i=1}^{l} Y_i=n$和$\sum_{i=1}^{l} p_i=1$，读者可从4.8节的多项式法则得此结论，即

$$P[Y_1=r_1,Y_2=r_2,\cdots,Y_l=r_l]=P(\boldsymbol{r};n,\boldsymbol{p})=\frac{n!}{r_1!r_2!\cdots r_l!}p_1^{r_1}p_2^{r_2}\cdots p_l^{r_l}$$

$$\approx \frac{\exp\left(-\frac{1}{2}\left[\sum_{i=1}^{l}\left(\frac{r_i-np_i}{\sqrt{np_i}}\right)^2\right]\right)}{\sqrt{(2\pi n)^{l-1}p_1p_2\cdots p_l}}, \qquad n>>1 \tag{7.4-1}$$

第j次试验的概率密度函数为$P_j=\prod_{i=1}^{l} p_i^{x_{ij}}$，其中$\sum_{i=1}^{l} x_{ij}=1$，$\sum_{i=1}^{l} p_i=1$，$x_{ij}$约束为0或1。$n$次重复试验的似然函数为$L(\boldsymbol{p})\triangleq L(p_1,\ \cdots,\ p_l)=\prod_{i=1}^{l} p_i^{X_{i1}}\prod_{i=1}^{l} p_i^{X_{i2}}\cdots\prod_{i=1}^{l} p_i^{X_{in}}=\prod_{i=1}^{l} p_i^{Y_i}$。在$H_1: p_i=p_{0i}$，$i=1,\ \cdots,\ l$假设下，似然函数的局部最大$L_{\mathrm{LM}}$为$L(\boldsymbol{p}_0)=\prod_{i=1}^{l} p_{0i}^{X_{i1}}\prod_{i=1}^{l} p_{0i}^{X_{i2}}\cdots\prod_{i=1}^{l} p_{0i}^{X_{in}}=\prod_{i=1}^{l} p_{0i}^{Y_i}$，似然函数的全局最大，分别为对$p_i$，$i=1,\ \cdots,\ l$微分获得，同时有$\sum_{i=1}^{l} p_i=1$。根据结果$\hat{p}_i=Y_i/n$，$i=1,\cdots,\ l$，有$L_{\mathrm{GM}}=L(Y_1/n,\ Y_2/n,\ \cdots,\ Y_l/n)=\prod_{i=1}^{l}(Y_i/n)^{Y_i}$，最后，应用$\sum_{i=1}^{l} Y_i=n$，可得广义似然比为

$$\Lambda=n^n\prod_{i=1}^{l}\left(\frac{p_{0i}}{Y_i}\right)^{Y_i} \tag{7.4-2}$$

判决域为$0<\Lambda<\lambda_c$。为在特定显著性水平上计算判决域，需要Λ的分布，然而，Λ在H_1假设下对于任意n值的准确分布难以得到。在参考文献[7-1]中证明了在大数据量假定下，$-2\ln\Lambda$近似于χ^2_{l-1}。

我们在此考虑另一种方法。式(7.4-1)可见在大数据量假定下，Y_i，$i=1,\ \cdots,\ l$近似为正态随机变量$N(np_i,\ np_i)$，$i=1,\ \cdots,\ l$，$U_i\triangleq\frac{Y_i-np_i}{\sqrt{np_i}}$，$i=1,\ \cdots,\ l$近似为标准正态。检验统计量

$$V\triangleq\sum_{i=1}^{l}\left(\frac{Y_i-np_{0i}}{\sqrt{np_{0i}}}\right)^2 \tag{7.4-3}$$

被称为皮尔逊检验统计量，基于V的大小判决假设的真假检验称为皮尔逊检验或开方检验(参

见参考文献[7-16]至参考文献[7-20])。皮尔逊检验统计量形式为l自由度的χ^2随机变量，但实际为$l-1$自由度，因为当Y_1, Y_2, …, Y_{l-1}给定时，$Y_l = n - \sum_{i=1}^{l-1} Y_i$随之确定。若$Y_i$来自于概率为$p_{0i}$, $i=1$, …, l的集合，我们期望V的实现是很小的。然而，若Y_i来自于概率为p_i, $i=1$, …, l的集合，其中至少两个p_i与对应的p_{0i}相差很大，我们期望V的实现是很大的。可以通过在H_1和H_2时计算$E[V]$来说明，在H_1时计算得到$E[V|H_1]=l-1$(参见习题7.24)，然而当n取值很大(参见习题7.25)，在H_2时$E[V|H_2]$为

$$E[V|H_2] \approx \sum_{i=1}^{l} (p_{0i})^{-1} n(p_{1i}-p_{0i})^2 \tag{7.4-4}$$

显然当有某些p_{1i}与p_{01}不同时，$E[V|H_2]$比$l-1$要大得多。准确的计算证明$E[V|H_2]$绝不会小于$l-1$。

回到式(7.4-3)的检验统计量，即

$$V \triangleq \sum_{i=1}^{l} \left(\frac{Y_i - np_{0i}}{\sqrt{np_{0i}}}\right)^2$$

注意到在H_1假设下为χ^2_{l-1}。为计算在显著性α时确定判决域$\{V>c\}$的常量c，我们计算$\int_c^{\infty} f_{\chi^2}(x; l-1)\mathrm{d}x = \alpha$，或等效计算$1-\alpha = F_{\chi^2}(c; l-1)$，可得$c = x_{1-\alpha}$，即$\chi^2_{l-1}$的$1-\alpha$百分点，则判决为：若$V<x_{1-\alpha}$，$H_1$为真，否则为假。

例 7.4-1 (掷硬币的公正性)我们在显著性$\alpha=0.05$时检验$p_{01}=P[\text{正面}]=0.5=p_{02}=P[\text{背面}]$的假设$H_1$，投掷硬币100次观测到61次正面，39次背面，则根据

$$V \triangleq \sum_{i=1}^{l} \left(\frac{Y_i - np_{0i}}{\sqrt{np_{0i}}}\right)^2$$

可得$V' = \frac{1}{0.5\times100}[61-50]^2 + \frac{1}{0.5\times100}[39-50]^2 = 4.84$。根据$0.95=F_{\chi^2}(x_{0.95}; 1)$计算判决值可得$x_{0.95}=3.84$，由于$V'=4.84>3.84$，我们判决掷硬币是公平的假设为假。

例 7.4-2 (掷骰子的公正性)我们在显著性$\alpha=0.05$时检验六面骰子是公正的假设。令Y_i, $i=1$, …, 6表示第i面的次数，投掷1000次观测到$Y_1'=152$, $Y_2'=175$, $Y_3'=165$, $Y_4'=180$, $Y_5'=159$, $Y_6'=171$，则

$$V' = \frac{1}{167}\Big[(167-152)^2 + (167-175)^2 + (167-165)^2 + (167-180)^2 + (167-159)^2 + (167-171)^2\Big] = 3.25$$

自由度为5，求解$0.95=F_{\chi^2_5}(x_{0.95})$得到$x_{0.95}=11.1$，由于$3.25<11.1$，我们判决掷骰子是公正的假设为真。

例 7.4-3 (正态性检验)我们希望确定数据是否来自标准正态$N(0, 1)$集合。令H_1表示X是标准正态$N(0, 1)$分布假设，H_2表示不是$N(0, 1)$的备选假设。累积正态分布$\{p_{0i}\}$如下：

$p_{01} \triangleq F_{\mathrm{SN}}(-2.0) = 0.023; p_{02} \triangleq F_{\mathrm{SN}}(-1.5) - F_{\mathrm{SN}}(-2.0) = 0.044; p_{03} \triangleq F_{\mathrm{SN}}(-1.0) - F_{\mathrm{SN}}(-1.5) = 0.092; p_{04} \triangleq F_{\mathrm{SN}}(-0.5) - F_{\mathrm{SN}}(-1.0) = 0.145; p_{05} \triangleq F_{\mathrm{SN}}(0) - F_{\mathrm{SN}}(-0.5) = 0.1915; p_{06} \triangleq F_{\mathrm{SN}}(0.5) - F_{\mathrm{SN}}(0) = 0.1915; p_{07} \triangleq F_{\mathrm{SN}}(1.0) - F_{\mathrm{SN}}(0.5) = 0.15; p_{08} \triangleq F_{\mathrm{SN}}(1.5) - F_{\mathrm{SN}}(1.0) = 0.092; p_{09} \triangleq F_{\mathrm{SN}}(2.0) - F_{\mathrm{SN}}(1.5) = 0.044; p_{010} \triangleq F_{\mathrm{SN}}(\infty) - F_{\mathrm{SN}}(2) = 0.023$

在1000次观测中得到如下现实：

在区间$(-\infty, -2]$：$Y_1'=19$
在区间$(-2, -1.5]$：$Y_2'=42$
在区间$(-1.5, -1]$：$Y_3'=96$
在区间$(-1, -0.5]$：$Y_4'=135$
在区间$(-0.5, 0]$：$Y_5'=202$
在区间$(0, 0.5]$：$Y_6'=193$
在区间$(0.5, 1]$：$Y_7'=155$
在区间$(1, 1.5]$：$Y_8'=72$
在区间$(1.5, 2]$：$Y_9'=53$
在区间$(2, \infty]$：$Y_{10}'=33$

应用 $V \triangleq \sum_{i=1}^{10}\left(\frac{Y_i - 1000p_{0i}}{\sqrt{1000p_{0i}}}\right)^2$ 作为检验统计量，其在 H_1 时 V 为 χ_9^2。对于给定数据计算得到 $V' = 12.9$，由于 $x_{0.95} = 16.92$，12.9 小于 16.92，我们判决数据来自正态分布的假设为真。

可以应用皮尔逊检验统计量来检验两个未知概率是否相同，即使没有其他如均值、方差等先验信息。例如，我们用印刷机测试两种品牌印刷纸：品牌A在150次试验中卡纸6次，品牌B在550次试验中卡纸25次，对印刷机而言，品牌A和品牌B是否一样卡纸呢？有两种语音识别程序可购买，假定同一个说话者，我们发现语音程序SR1在250个单词中出错61次，SR2在110个单词中出错30次，两个程序是否效率相同呢？在概率论框架内我们将此建模如下：考虑两类事件 E_1 和 E_2 的发生，是否满足 $P[E_1]=P[E_2]$？定义 Z_1 为 m 次试验中我们观察 E_1 的发生次数，Z_2 为后续 n 次试验中我们观察 E_2 的发生次数，令 $p_1 \triangleq P[E_1]$ 及 $p_2 \triangleq P[E_2]$，当 $m \gg 1$，$n \gg 1$ 时，根据中心极限定理有 $Z_1: N(mp_1, mp_1q_1)$ 及 $Z_2: N(np_2, np_2q_2)$。定义标准随机变量 $Y_1 \triangleq Z_1/m: N(p_2, p_1q_1/m)$ 及 $Y_2 \triangleq Z_2/n: N(p_2, p_1q_1/n)$ 并考虑随机变量 $Y \triangleq Y_1 - Y_2$，由于 Y_1 和 Y_2 为独立的（回想 Y_1 是从前面 m 次试验观察得到，Y_2 是从后续 n 次试验观察得到），可知 Y 为均值为 $p_1 - p_2$、方差为 $\sigma_Y^2 = (np_1q_1 + mp_2q_2)/mn$ 的正态随机变量。令 H_1 为 $p_1 = p_2$ 的假设，H_2 为 $p_1 \neq p_2$ 的备选假设，显然在 H_1 条件下 $Y: N(0, p_1q_1(m+n)/nm)$。此问题的皮尔逊检验统计量为

$$V = \left(\frac{Y - (p_1 - p_2)}{\sigma_Y}\right)^2$$

服从 χ_1^2 分布。在显著性 α 时检验，计算满足 $1-\alpha = F_{\chi^2}(x_{1-\alpha}; 1)$ 的百分点 $x_{1-\alpha}$，若 $V < x_{1-\alpha}$，判决假设为真，否则判决假设为假。

此问题的难点在于由于 p_1 和 p_2 未知，从而 σ_Y 是未知的，一个解决方案是用基于观测的 σ_Y 的估计量来替代 σ_Y。在 H_1 时，$p_1 = p_2 \triangleq p$，p 的最小方差无偏估计量为 $\hat{p} = (Z_1 + Z_2)/(m+n)$，可知在 H_1 时，有 $\hat{\sigma}_Y = \sqrt{\hat{p}\hat{q}(m+n)/mn}$，其中 $\hat{q} = 1 - \hat{p}$。我们用两个例子来阐述。

例7.4-4 （不同地区的投票模式）在一个大州的选举中，投票后民意测验显示州北部农业地区211选举人中167人投票给共和党，南部城市地区499选举人中216人投票给共和党，是否可以假定北部地区投票给共和党的概率 p_1 与南部地区投票给共和党的概率 p_2 相同？

解 在H_1时，$p_1=p_2 \triangleq p$，在H_2时，$p_1 \neq p_2$。在H_1时，计算可得$\hat{p}'=388/710=0.54$，$\hat{q}'=0.46$，$\hat{\sigma}_Y'=0.041$，$Y_1'=167/211=0.79$，$Y_2'=216/499=0.43$及$Y' \triangleq Y_1'-Y_2'=0.36$，因而$V'=(0.36/0.041)^2 \approx 77$，在显著性水平$\alpha=0.05$时，可得$x_{0.95}=3.84$。由于$77>3.84$，假设显然不成立。

例 7.4-5 (科学数据的解释)在对恒河猴是否能区分不同的声音并与某种含义建立联系，包括口头语言，做了下述实验。一个恒河猴放入无回声的房间(外部隔音)，房间内有一个计算机控制方向的喇叭，随机地发出两种信号中的一种S_1，一种恒河猴在其自然居住地可能听到的声音S_2，一种口头语音。如果猴子当听到声音发出时，将头转向喇叭，说明猴子对声音有反应。如果发出S_2，可能意味着猴子对声音好奇或感兴趣，可能用此声音作为命令进行训练。但是，如果猴子对此声音没有反应，意味着猴子认为它没有意义。从研究者角度，理想情况是当发出S_1时，猴子从不转头，当发出S_2时，猴子总是转头，如此研究者就可以写一篇关于恒河猴认知能力的学术论文并因此出名①。我们忽略猴子的头是否转了足够的角度而可称为“转头”这个令人困惑的问题②。

在267次发出“自然居住地”声音中，猴子转头112次；在289次发出口头语音中，恒河猴转头173次。令p_1表示发出“自然居住地”声音时猴子转头的概率，令p_2表示发出口头语音时猴子转头的概率，在H_1时，$p_1=p_2 \triangleq p$，在H_2时，$p_1 \neq p_2$。我们是否可以假定猴子对两种声音的反应没有区别，即H_1为真呢?

解 在H_1时，有$\hat{p}'=0.51$，$\hat{q}'=0.49$，$\hat{\sigma}'=0.0424$，$Y_1'=112/267=0.42$，$Y_2'=173/289=0.6$及$Y' \triangleq Y_1'-Y_2'=0.18$，因而$V'=(0.18/0.0424)^2=18$。在0.05的显著性水平时有$x_{0.95}=3.84$，因此假设显然不成立。

7.5 排序，百分点及排名

为方便读者，我们将6.8节的部分内容复述一下。取累积分布函数为$F_X(x)$的一类随机变量X(有时称为集合)的n个独立同分布观测$X_1, X_2, \cdots, X_n$，样本的联合概率密度函数为$f_X(x_1)\times\cdots\times f_X(x_n)$，$-\infty<x_i<\infty$，$i=1,\cdots,n$。我们将$X_i$，$i=1,\cdots,n$按大小(符号数值)排序后得到排序的样本$Y_1, Y_2, \cdots, Y_n$，即$-\infty<Y_1<Y_2<\cdots<Y_n<\infty$，有时称为$X$观测的排序统计量。当排序后，序列3，-2，-9，4变成-9，-2，3，4，。若序列$X_1,\cdots,X_{20}$是由$X:N(0,1)$的n个观测生成，则$Y_1>0$的可能性不大，因为这需要其他19个Y_i，$i=2,\cdots,20$都大于零，即所有样本都在正态曲线的正侧，这种可能性为$(1/2)^{20}$。同样，$Y_{20}<0$也是不太可能的，因为这需要其他19个Y_i，$i=1,\cdots,19$都小于零。如5.3节所示，排序样本$Y_1, Y_2, \cdots, Y_n$的联合概率密度函数为$n!\, f_X(y_1)\times\cdots\times f_X(y_n)$，$-\infty<y_1<y_2<\cdots<y_n<\infty$或零。排序和排名是不同的，排名通常给排序单元指派一个值，例如，大多数人会将骨折的疼痛排序在感冒引起的喉咙疼痛前面，但若一个内科医生要病人将疼痛按0到10排名，骨折的疼痛可能排名第8或第9，而喉咙疼痛可能排名第3或第4。

---

① 这项研究已由一些大学完成，但结论在科学界引起争论。

② 一个类似于棒球中“检查摆幅”问题，裁判员必须判决击球手是否“完成击球后的弧形动作”还是“检查其摆幅”。

下面考虑百分点的概念，在本书前面很多地方讨论中已应用过此概念，我们将在此就其细节做简短阐述。假定一大群人的 IQ 分布为 $N(100, 100)$，即均值为 100，标准差为 10。显然，正态近似是在一定范围内容有意义的，因为没有人的 IQ 会是 1000 或是 −10。IQ 测试本身也是在一定范围内有意义的，对于特别聪明或严重智力低下的也难以给出精确的数值，有时将这两类人叫做“脱离 IQ 范围”。现阶段 IQ 测试依然作为解决问题能力的指标广泛应用，假定 IQ 测试结果表明小孩在应试者中排名是 93 百分点，由于其“有天分”而有资格参加项目，我们怎么定位该百分点呢？

定义（百分点）　给定累积密度函数 $F_X(x)$ 的随机变量 X，X 的 μ 百分点是满足 $F_X(x_u)=u$ 的 x_u 值。若连续函数 F_X 是处处连续且可微的，则 $x_u=F_X^{-1}(u)$，其中 F_X^{-1} 是累积密度函数 F_X 的逆函数，满足 $F_X^{-1}(F_X(x_u))=x_u$。标准正态累积密度函数及其逆函数如图 7.5-1 所示。

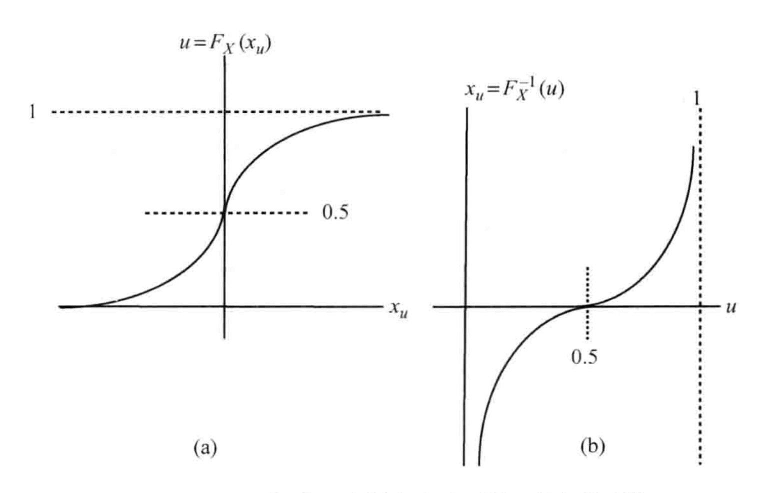


图 7.5-1　(a)标准正态累积密度函数；(b)逆函数

观测　在标准正态 $Z:N(0, 1)$ 的特定情况，我们用符号 z_u（表示 X 的 u 百分点。若 $X:N(\mu, \sigma^2)$，则 X 的 u 百分点 x_u 与 z_u 的关系满足

$$x_u = \mu + z_u\sigma \tag{7.5-1}$$

例 7.5-1　（x_u 与 z_u 的关系）证明 $x_u=\mu+z_u\sigma$。

解　我们有

$$\begin{aligned}F_X(x_u) = u &= \left(2\pi\sigma^2\right)^{-1/2}\int_{-\infty}^{x_u}\exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)\mathrm{d}x \\ &= (2\pi)^{-1/2}\int_{-\infty}^{(x_u-\mu)/\sigma}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}z \\ &\triangleq (2\pi)^{-1/2}\int_{-\infty}^{z_u}\exp\left(-\frac{1}{2}z^2\right)\mathrm{d}z\end{aligned}$$

最后一行是 $Z:N(0, 1)$ 的累积密度函数，因此 $x_u=\mu+z_u\sigma$。我们将此结论用于前面提到的 IQ 问题。根据数据有 $F_X(x_u)=0.93=F_Z(z_u)$，从 F_{SN} 表中可得 $z_u\approx1.48$，根据 $x_u=\mu+z_u\sigma=100+1.48(10)$，得到 IQ 的 93 百分点对应的 IQ 为 115。

估计百分点和中值时排序的作用①

简短回顾6.8节与百分点和中值相关的部分内容。

集合 X 的中值是满足 $F_X(x_{0.5})=0.5$ 的点 $x_{0.5}$，而 X 的均值，写成 μ_X，定义为 $\mu_X=\int_{-\infty}^{\infty}xf_X(x)\mathrm{d}x$。中值和均值并不需要一致，例如，在 $f_X(x)=\lambda e^{-\lambda x}u(x)$ 情形中，可得 $\mu_X=1/\lambda$ 及 $x_{0.5}=0.69/\lambda$。计算 X 的均值需要 $f_X(x)$，但这经常是不知道的。均值看起来是一个相对抽象的参数，而中值仅仅是一个将集合分成两部分的点，即集合一半在这个点之上，一半在它之下②。当 $f_X(x)$ 假定存在且其参数可获取或估计时称为参数情形。在典型的参数情形中，需要假定集合密度的形式，如正态分布，并希望估计分布中的某些未知参数，如均值 μ_X。对于 X 的给定 n 个独立同分布观测 $X_1, X_2, \cdots, X_n$，根据 $\hat{\mu}_X=n^{-1}\sum_{i=1}^{n}X_i$ 估计 μ_X，其对于很多集合而言是无偏且一致估计量。实际上其是均值估计函数 $\hat{\mu}_X$ 的简单形式，当 σ_X^2 有限且 n 很大(参见大数定理)时，有 $\hat{\mu}_X\to\mu_X$，这使得在很多应用中均值非常有用。估计已知或假定分布中的参量，以及包括已知或假定分布的假设检验等运算，称为参数统计。

集合的特性和参数的估计不依赖于集合分布的知识或形式的假定称为分布无关的，鲁棒的或非参量统计。仅仅基于观测而没有分布的假定的统计称为鲁棒的，这是因为根据观测得到的定理或结论不会因为分布形式的不同而改变。在参量情形中，均值和标准差在描述集合的中心和散布时是十分有用的，中值和范围在非参量情形中扮演相同的角色。为估计 $X_1, X_2, \cdots, X_n$ 的中值，我们根据排序统计量用样本中值估计量来估计 $x_{0.5}$

$$\begin{aligned}\hat{Y}_{0.5}&=Y_{k+1}\text{ 若 } n \text{ 为奇数，即 } n=2k+1\\&=0.5(Y_k+Y_{k+1})\text{ 若 } n \text{ 为偶数，即 } n=2k\end{aligned}\tag{7.5-2}$$

样本中值不是 $x_{0.5}$ 的无偏估计，但在 n 很大时近似无偏。非参量情形的散布用50%百分点范围描述，即 $\Delta x_{0.50}\triangleq x_{0.75}-x_{0.25}$，或用90百分点范围描述，即 $\Delta x_{0.90}\triangleq x_{0.95}-x_{0.05}$，或用其他合适的范围描述。

例7.5-2 (百分点的内插)用符号 $\alpha\sim\beta$ 表示 α 估计 β，有 $Y_3\sim x_{0.273}$，$Y_4\sim x_{0.364}$，应用内插公式，可得 $x_{0.3}$ 为

$$Y_4+\frac{(Y_4-Y_3)(0.3-4/11)}{1/11}\sim x_{0.3}$$

图7.5-2显示了排序样本间的内插。

下面讨论一个与百分点排序统计相关的基本结论。模型依然是累积概率密度为 $F_X(x)$ 的随机变量 X 的 n 个独立同分布观测 $X_1, X_2, \cdots, X_n$，应用标记 $P[X_i\leqslant x_u]\triangleq u$，下面考虑排序统计量 $Y_1<Y_2<\cdots<Y_n$ 及事件 $\{Y_k<x_u\}$，由于 Y_k 是 $\{X_i\}$ 排序中第 k 个单元，表明在 $\{X_i\}$ 中至少有 k 个值小于 x_u，可能有更多，但不会少。因为 $\{X_i\}$ 是独立同分布的，可以用二项式概率公式计算 $P[Y_k<x_u]$ 为

---

① 对6.8节内容熟悉的读者可跳过此部分。

② 中值“集合一半在其之上，一半在其之下”的说法不是很精确。读者应当意识到样本的中值通常与整个集合的中值是不同的。

$$P[Y_k < x_u] = P\left[\{X_i\}\text{中至少有}k\text{个值小于 }x_u\right]$$
$$= \sum_{i=k}^{n}\binom{n}{i}u^i(1-u)^{n-i} \tag{7.5-3}$$

下面考虑事件$\{Y_{k+r} > x_u\}$，由于Y_{k+r}是$\{X_i\}$排序中第$(k+r)$个单元，则$\{X_i\}$中至少有$n-(k+r)+1$个值大于x_u，等效地，$\{X_i\}$中有不超过$k+r-1$个值小于x_u。即

$$P[Y_{k+r} > x_u] = P[\{X_i\}\text{中有不超过}k+r-1\text{个值小于}x_u]$$
$$= \sum_{i=0}^{k+r-1}\binom{n}{i}u^i(1-u)^{n-i} \tag{7.5-4}$$

事件$\{Y_{k+r} > x_u\} \cup \{Y_k < x_u\}$的并集为事件$\{Y_k < x_u < Y_{k+r}\}$，其概率

$$P[Y_k < x_u < Y_{k+r}] = \sum_{i=k}^{k+r-1}\binom{n}{i}u^i(1-u)^{n-i} \tag{7.5-5}$$

与$f_X(x)$是相互独立的。式(7.5-5)给出的结论是非参量统计中的重要结论并有重要应用，例如下述集合的中值估计。

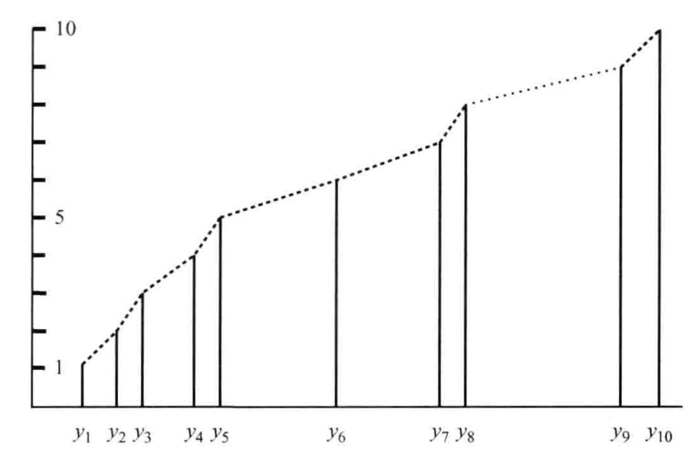


图 7.5-2　从 10 个排序的样本估计百分点范围显示为样本间线性内插。为获取估计的百分点，采用纵坐标值并乘以100/11，这样，对于初始近似，90百分点从y_{10}估计，而9百分点从y_1估计，y_8-y_2覆盖了50百分点范围

例 7.5-3　（覆盖中值的95%置信度需要多大的样本数?）我们寻求随机区间$[Y_1, Y_n]$的端点Y_1，Y_n，满足事件$\{Y_1 < x_{0.5} < Y_n\}$发生的概率是 0.95。此处$Y_1 \triangleq \min(X_1, X_2, \cdots, X_n)$，$Y_n \triangleq \max(X_1, X_2, \cdots, X_n)$，则实际上$n$应该取多大呢?

解　通过计算

$$P[Y_1 < x_{0.5} < Y_n] = \sum_{i=1}^{n-1}\binom{n}{i}(1/2)^n \approx 0.95$$

可得当$n=5$，$P[Y_1 < x_{0.5} < Y_5] \approx 0.94$，随机区间$[Y_1, Y_n]$覆盖50%的百分点的概率随$n$的变化规律在图 7.5-3 中示出。

例 7.5-4　（覆盖$x_{0.33}$的最可能排序对）有一组排序统计量$\{Y_1, Y_2, \cdots, Y_n\}$，希望找到一个排序对$\{Y_i, Y_{i+1}, i=1, \cdots, n-1\}$覆盖 33.33 百分点的概率最大，33.33 百分点由 1/3 =

$F_X(x_{0.33})$给出。我们假定 $n=10$，根据式(7.5-5)可得

$$P[Y_k < x_{0.33} < Y_{k+1}] = \frac{10!}{k!(10-k)!}(1/3)^k(2/3)^{10-k}, k=1,\cdots,9$$

结果绘在图7.5-4中。显然间隔$[Y_3, Y_4]$覆盖$x_{0.33}$的可能性最大，事件$\{Y_3 < x_{0.33} < Y_4\}$的概率为0.26。

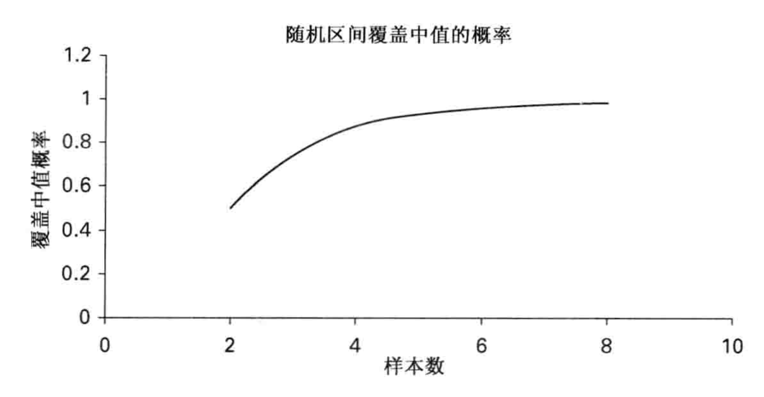


图7.5-3　事件$\{Y_1 < x_{0.5} < Y_n\}$覆盖中值的概率与 n 值的关系

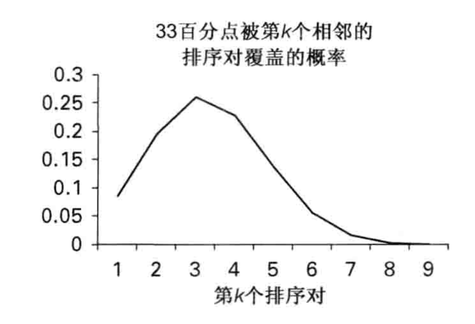


图7.5-4　在所有成对区间$[Y_k, Y_{k+1}]$中，区间$[Y_3, Y_4]$覆盖$x_{0.33}$的可能性最大，此处 $n=10$

例7.5-5　(对于二项式分布，中值和均值不同)下面我们做一些常规观测，对于二项式情形，均值和中值是不同的。例如，当 $p=1/2$，$n=4$ 时，均值为2，其中值将会位于1和2之间。然而，当 n 取值很大，均值和中值相互接近，中值可用均值估计得出，事实上，均值和中值的误差正比于$(p(1-p))^n$，当 $n\to\infty$ 时可以任意小。

当 n 很大时中值的置信区间

若 n 足够大，以至于二项式分布可以用正态分布近似，我们有

$$P[\alpha \leqslant S_n \leqslant \beta] \approx \frac{1}{\sqrt{2\pi}}\int_{\alpha_n}^{\beta_n} \exp\left[-\frac{1}{2}y^2\right]\mathrm{d}y$$

其中

$$P[\alpha \leqslant S_n \leqslant \beta] = \sum_{i=\alpha}^{\beta}\binom{n}{i}p^i(1-p)^{n-i}$$
$$\alpha_n \triangleq \frac{\alpha - np - 0.5}{\sqrt{np(1-p)}},\ \beta_n \triangleq \frac{\beta - np + 0.5}{\sqrt{np(1-p)}} \tag{7.5-6}$$

应用此结果于就近的问题，有

$$P[Y_r < x_{0.5} < Y_{n-r+1}] = \sum_{i=r}^{n-r}\binom{n}{i}(1/2)^n \tag{7.5-7}$$

在应用中中值定义为 $u=F_X(x_{0.5})=1/2$。下标的选择确保置信区间于从底部开始的第 r 个位置计算，即1，2，3，…，r，结束于从顶部开始往回计算 r 个观测位置。举例来说，如果 $n=10$ 时95% 置信区间得到 $r=3$，置信区间从第3个观测开始，结束于第8个观测，两个点均分别是从底部或顶部开始的第3个位置，即1，2，3(Y_3)和10，9，8(Y_8)，结果可表示为 $P[Y_3<x_{0.5}<Y_8]=0.95$。

从二项式求和公式[参见式(7.5-7)]注意到，其均值为 $n/2$，标准差为 $\sqrt{n}/2$，而二项式求和公式的95% 置信区间正态近似为

$$\sum_{i=r}^{n-r}\binom{n}{i}(1/2)^n \approx \frac{1}{\sqrt{2\pi}}\int_{\alpha_n}^{\beta_n}\exp[-\frac{1}{2}x^2]\mathrm{d}x = 0.95$$

由标准正态分布函数 $F_{SN}(x)$ 表可得 $\alpha_n=-1.96$，$\beta_n=1.96$。根据式(7.5-6)，有

$$1.96 = \frac{n-r-n/2+0.5}{\sqrt{n}/2}$$

$$-1.96 = \frac{r-n/2-0.5}{\sqrt{n}/2}$$

可得 $r=(n/2)-1.96\sqrt{n}/2+0.5$。如果 r 不是整数，则用 $\lfloor r\rfloor$ 代替 r，$\lfloor r\rfloor$ 为小于等于 r 的最大整数。

例7.5-6 （$n=20$ 时中值的95% 置信区间）随机变量 X 的20个观测为 $\{X_i, i=1, \cdots, 20\}$，根据其大小排序为 $Y_1<Y_2<\cdots<Y_n$，由 $r=(n/2)-1.96\sqrt{n}/2+0.5$ 可得 $r=6.12$ 和 $\lfloor r\rfloor=6$，则有 $P[Y_6<x_{0.5}<Y_{15}]\geqslant 0.95$。

分布无关假设检验：采用游程法检验两个集合是否相同

通常假设检验采用非参数统计量比参数情形要复杂难懂，因为检验统计量的分布难以计算，然而，当样本数很大时，比如大于10，我们可以应用正态近似来计算判决域。

通过考虑下述简单情况来介绍游程法。我们得到累积密度函数为 $F_X(x)$ 的随机变量 X(集合) n_1 个观测 $\{X_i^{(1)}, i=1, \cdots, n_1\}$，将其按大小排序后得到样本 $\{Y_i, i=1, \cdots, n_1\}$，我们再得到同一随机变量 X 的 n_2 个观测 $\{X_i^{(2)}, i=1, \cdots, n_2\}$，将样本按大小排序后得到 $\{Z_i, i=1, \cdots, n_2\}$。然后将两组未排序的样本合并后按大小排序，则排序后序列可能为 Z_1，Z_2，Y_1，Z_3，Y_2，…，Z_{n_2}，Y_{n_1-1}，Y_{n_1}，其中 $Z_1<Z_2<Y_1<Z_3<Y_2<\cdots<Z_{n_2}<Y_{n_1-1}<Y_{n_1}$。定义一个游程为被其他类的字母限制的同类字母序列或整个序列的开始/结束，因此 Z_1，Z_2 为首个游程，其长度为2，下个游程为 Y_1，其长度为1，以此类推，最后一个游程为 $Y_{n_1-1}Y_{n1}$，长度为2。计算游程总数并称为 D，注意到 D 为随机变量。由于两个样本集来自同一集合，可以期望 Y 集和 Z 集会混合得相当彻底，从而 D 会很大。然而，若 Y 集和 Z 集来自不同集合，D 很有可能会显著减小。例如假定有两个集合，即概率密度函数为 $f_{X(1)}(x)=\mathrm{rect}(x)$ 的集合 $X^{(1)}$ 和概率密度函数为 $f_{X(2)}(x)=\mathrm{rect}(x-2)$ 的集合 $X^{(2)}$，若 $\{Y_i, i=1, \cdots, n\}$ 表示 $X^{(1)}$ 集合的排序，$\{Z_i, i=1, \cdots, n\}$ 表示 $X^{(2)}$ 集合的排序，则混合序列的排序样本为 $Y_1Y_2\cdots Y_nZ_1Z_2\cdots Z_n$，由于其概率密度函数没有重合，有 $D'=2$。

例7.5-7　(均值相同和不同时集合的 D 的现实)利用互联网上的随机数发生器 RANDOM. ORG 生成两个正态随机数集，每个集合10个数

$$N(0,1) \to \{x^{(1)}:\ -0.19,\ 0.99,\ -1.1,\ -1.0,\ -1.3,\ -0.53,\ -0.25,\ 0.75,\ -0.25,\ 0.75\ \}$$
$$N(0,1) \to \{x^{(2)}:\ 0.68,\ -1.2,\ 0.28,\ 0.61,\ -1.2,\ -1.5,\ 2.1,\ -0.10,\ -0.87,\ 0.80\ \}$$

将序列 $x^{(1)}$ 和 $x^{(2)}$ 分别按大小排列得到排序后序列 $y_1y_2\cdots y_{10}$ 及 $z_1z_2\cdots z_{10}$，其中 $y_1=-1.3$，$y_{10}=0.99$，$z_1=-1.5$ 及 $z_{10}=2.1$，将两个序列混合成一个序列并按大小排序，得到序列 $z_1y_1z_2z_3y_2y_3z_4y_4y_5y_6y_7z_5z_6z_7z_8y_8y_9z_9y_{10}z_{10}$，可得 $D'=11$。

重复实验，从标准正态分布 $N(0,1)$ 取出10个随机数，从 $N(1,1)$ 取出另外10个，数据按不同位置存放，结果为

$$N(0,1) \to \{x^{(1)}:\ -0.079,\ 1.3,\ -0.15,\ 1.2,\ 0.75,\ -1.2,\ -0.11,\ -0.84,\ 0.35,\ 0.55\ \}$$
$$N(1,1) \to \{x^{(2)}:\ 1.2,\ 0.056,\ 0.3,\ -0.77,\ 0.95,\ 1.1,\ 0.095,\ -0.43,\ 1.1,\ 1.3\}$$

此处排序的 y 序列与 $N(0,1)$ 相关，排序的 z 序列与 $N(1,1)$ 相关，将数据合并成单一序列后再按大小排序，得到序列

$$y_1y_2z_1z_2y_3y_4y_5z_3z_4z_5y_6y_7y_8z_6z_7z_8z_9y_9y_{10}z_{10}$$

可得 $D'=8$，比单独 $N(0,1)$ 情况小27%。这个例子说明随机变量 D 可以作为两个集合是否相同的假设的检验统计量。若 D 足够大，比如说 $D>d_0$，可以得出两个样本集出自同一集合的结论，否则就不是出自同一集合。下面讨论 d_0 的选择。

我们采用假设检验的原理来检验两个样本集是否出自同一集合。有两个样本集：$\{X_i^{(1)}, i=1,\cdots,n_1\}$ 和 $\{X_i^{(2)}, i=1,\cdots,n_2\}$。零假设 H_1 为两个样本集来自同一集合，备选假设 H_2 为两个样本集不是来自同一集合，可能更精确的描述是：没有足够的证据表明它们来自同一集合。检验是基于观测检验统计量 D，若事件 $\{D>d_0\}$ 发生，则两个样本集混合充分从而来自同一集合，若事件 $\{D\leqslant d_0\}$ 发生，则表明数据不支持 H_1 成立。若 $\alpha \triangleq P[H_1 \text{为假} \mid H_1 \text{为真}]$ 表示显著性水平，则 $\alpha = P[D\leqslant d_0 \mid H_1 \text{为真}] = \sum_{\text{所有}d\leqslant d_0} P_D(d;n_1,n_2)$，其中 $P_D(d;n_1,n_2)$ 为在长度为 n_2 和 n_2 交织序列中观测到 d 游程的概率。

计算 $P_D(d;n_1,n_2)$ 需要较复杂的计算过程，此处我们仅仅给出最终结果。定义

$$C_m^n \triangleq \binom{n}{m}$$

在零假设时可得

$$P_D(d;n_1,n_2)=\begin{cases} 2C_{(d/2)-1}^{n_1-1}C_{(d/2)-1}^{n_2-1}/C_{n_1}^{n_1+n_2}, d\text{ 为偶数} \\ (C_{(d-1)/2}^{n_1-1}C_{(d-3)/2}^{n_2-1}+C_{(d-3)/2}^{n_1-1}C_{(d-1)/2}^{n_2-1})/C_{n_1}^{n_1+n_2}, d\text{ 为奇数} \end{cases}$$

这个笨拙的方程不能为分析目的给出更多信息，需要计算机来估计 α。然而，这显示了当 $n_1\geqslant 10$，$n_2\geqslant 10$，D 的分布十分接近正态累积密度函数，其均值和方差分别如下：

$$\mu_D \approx \frac{2n_1n_2}{n_1+n_2},\ \sigma_D^2 \approx 4(n_1+n_2)\left(\frac{n_1}{n_1+n_2}\right)^2\left(\frac{n_2}{n_1+n_2}\right)^2$$

我们可将 $\alpha = P[D\leqslant d_0 \mid H_1 \text{为真}] = \sum_{\text{所有}d\leqslant d_0} P_D(d;n_1,n_2)$ 近似为

$$\alpha = \sum_{\text{所有}d\leqslant d_0} P(d) \approx \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z_\alpha}\exp\left(-\frac{1}{2}x^2\right)\mathrm{d}x,\ z_\alpha \triangleq \frac{d_0-\mu_D}{\sigma_D}$$

例7.5-8 （相同集合的游程检验）我们应用RANDOM. ORG从集合$N(1, 1)$得到两个10个数据集，将其排序后

$$N(1,1) \to \{y^{(1)}: -1.4, -0.33, 0.40, 0.44, 0.70, 0.74, 1.3, 1.3, 1.7, 2.4\}$$
$$N(1,1) \to \{y^{(2)}: -0.67, -0.21, 0.38, 0.38, 0.51, 0.71, 1.4, 1.5, 2.0, 2.9\}$$

为测量目的，将两个序列合并成单一序列并按大小排序，可得现实$D'_{\mathrm{cal}}=12$。下面需要一个来自“未知”分布的随机数据集并将其按大小排序为

$$\{y^{(3)}: -3.8, -2.5, -0.13, 2.2, 2.8, 3.0, 3.8, 4.6, 5.5, 5.8\}$$

将此序列与$\{y^{(1)}\}$序列合并后排序计算游程，可得$D'=6$。我们希望在显著性水平0.05下检验$\{y^{(1)}\}$和$\{y^{(3)}\}$来自同一集合的假设，求解

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z_{0.05}} \exp\left(-\frac{1}{2}x^2\right)\mathrm{d}x = 0.05$$

可得$z_{0.05}=-1.65$。对于给定样本数，可得$\mu_D=10$，$\sigma_D=\sqrt{5}$，因而$d_0=\sigma_D z_{0.05}+\mu_D=6.3$，由于$D'<d_0$（勉强），我们判决$\{y^{(3)}\}$也来自$N(1, 1)$集合的假设不成立。事实上，$\{y^{(3)}\}$序列来自集合$N(1, 3)$。

相同两集合的排名检验

检验两个集合相同的另一程序称为排名检验。假定两个连续集合X和Y，其分布分别为$F_X(x)$和$F_Y(y)$，我们希望检验假设H_1：$F_X=F_Y$对备选H_2：$F_X\neq F_Y$。从X中取n_1个样本，从Y中取n_2个样本，将其混合并按大小排序，给序列中每个单元指定一个值表示其升序位置，例如，事件$X_1<Y_1<X_2<X_3<Y_2<Y_3<Y_4$指定为

$$\begin{matrix} X_1 & Y_1 & X_2 & X_3 & Y_2 & Y_3 & Y_4 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix}$$

指派的值为其单元的排名，Y序列排名为2，5，6和7，此处$n_1=3$，$n_2=4$。序列最后单元的排名为n_1+n_2，首单元排名为1。可证明随机变量

$$T \triangleq \sum_{Y\text{序列}} \text{排名}$$

是检验$F_X(x)=F_Y(x)$（对所有x）假设的合适的检验统计量，若T太大或太小，假设不成立。为在显著性水平α时检验假设，需要T在零假设下的分布，可以证明[7-22]~[7-24])，当$n_1>7$，$n_2>7$（理想情况是越大越好）时，T近似为$N(\mu_T, \sigma_T^2)$的分布，其中$\mu_T=n_2(n_1+n_2+1)/2$，$\sigma_T^2=n_1n_2(n_1+n_2+1)/12$。在上述例题中，可得$\mu_T=16$，$\sigma_T^2=8$。

例7.5-9 （相同集合的游程检验）应用例7.5-8中的$\{y^{(1)}\}$和$\{y^{(3)}\}$序列，将其合并后按升序给每个单元指定排名。对于$\{y^{(3)}\}$序列中的单元，排名为1，2，5，13，15，16，17，18，19和20，其和为126，$\mu_T=105$，$\sigma_T=13.23$。假设为两个序列来自相同集合。在显著性水平$\alpha=0.05$时，求解$F_T(x_{0.025})=0.025$得到$x_{0.025}=-1.96$，因此判决域为$\{T>131\}\cup\{T<79\}$，我们判决两个序列来自同一集合的假设为真（错误的）。在显著性水平$\alpha=0.1$时，判决域为$\{T>127\}\cup\{T<87\}$，在$\alpha=0.1$上面一点，假设将不成立。

小结

假设检验是统计学的一个主要内容，是在随机（或概率）环境下做出判决。在本章开始，将自己当成一名医生，要做出是否对其病人实施手术的困难决定，应用所有的先验信息并

寻求平均风险最小，我们推导出贝叶斯判决准则，在可用信息是概率类型而不是确定型时，贝叶斯判决准则是做出判决最合理的方法(有争议)，贝叶斯判决准则归结为似然比检验(LRT)。

贝叶斯检验中的先验概率并不总是可用的，比如似然比检验中假设判决门限是由指定错误概率α而不是平均风险最小化决定时，错误概率是当假设为真而基于观测的判决为假的概率。在检验简单假设对简单备选假设情形，纽曼－皮尔逊理论确保似然比检验是最佳的，即最大势检验，这意味着当假设为真时，判决假设不成立的概率最小。

在很多情况下，由于假设或备选或两者在样本空间包含很多结果，简单假设对简单备选假设的检验变得不可能，此时可用广义似然比检验(GLRT)。我们通过一系列例题介绍广义似然比检验，其中有 F 检验、t 检验及皮尔逊开方检验等经典检验。

然后考虑排序、百分点和排名，介绍了这些工具能够用于分布无关(有时称为鲁棒)统计，这些内容我们是通过游程检验和排名检验的假设检验例题来介绍的。

习题

7.1 证明式(7.1-6)。

7.2 考虑例7.1-1，令先验概率为 $P_1=0.9$，$P_2=0.1$，其将如何影响贝叶斯判决准则?

7.3 假定一个正态集合 $X:N(\mu, 1)$，对 X 的独立同分布观测序列为 $\{X_i: i=1, \cdots, n\}$，在显著性水平0.05时，计算 $H_1: \mu=\mu_1$ 对备选 $H_1: \mu>\mu_1$ 假设检验的判决域。

7.4 证明似然比检验的势 P 由 $P=P[H_1\text{为假}|H_2\text{为真}]$ 给出。

7.5 例7.2-2中，为什么不需要用中心极限定理来证明 $\hat{\mu}_X(n)$ 是正态分布的?

7.6 掷100次硬币，观察到 $50+k$ 次正面和 $50-k$ 次背面，当显著性 $\alpha=0.05$ 时，为确保掷硬币是公正的这个假设成立，k 的最大值是多少? 当 $\alpha=0.01$ 重复分析。

7.7 一个消费者在冰点温度环境下考虑在蓄电池折扣店购买汽车蓄电池(DBW)，感兴趣的蓄电池是两种渠道进口的，A和B，两种渠道并不是采用同一质量控制标准的，好的(A)可以在冰点温度时启动汽车的概率为90%，而差的(B)在此环境中启动汽车的概率为50%，两种渠道的蓄电池数目相同，且从外部特征不能区别。蓄电池销售商允许消费者在购买前用测试蓄电池在冰点环境下启动一次汽车。

我们需要从假设检验角度来面对消费者的困境。令假设为 H_1：蓄电池启动概率 $p_1=0.9$ 对备选假设 H_2：蓄电池启动概率 $p_2=0.5$，有两种操作：a_1(购买蓄电池)和 a_2(不购买蓄电池)，代价函数用美元表示：$l(a_1, p_1)=0$；$l(a_1, p_2)=40$(花在质量差的蓄电池上的钱)；$l(a_2, p_1)=10$(错过好蓄电池而在别处至少花费10美元)；$l(a_2, p_2)=0$。定义随机变量 X 为

$$X \triangleq \begin{cases} 1, \text{蓄电池可以启动汽车} \\ 0, \text{蓄电池无法启动汽车} \end{cases}$$

(a)确定四种可能判决函数(d_i, $i=1, \cdots, 4$)；

(b)计算每种判决函数的风险($R(d_i; p_j)$, $i=1, \cdots, 4$; $j=1, 2$)；

(c)在笛卡儿坐标系上绘出风险函数，横坐标为 $R(d; p_1)$，纵坐标为 $R(d; p_2)$。从图上确定哪种函数是由至少其他一种判决函数控制(更差)，因而是不可接受的(不值得考虑)。

(d)假定已知B类蓄电池数是A类的两倍，这会如何影响你的判决?

7.8 对于一个给定的简单假设及其备选假设问题，其判决风险集合 $[R(d; \theta_1), R(d; \theta_2)]$，当 $0\leqslant R(d; \theta_1)\leqslant 1$, $0\leqslant R(d; \theta_2)\leqslant 1$ 时，在考虑所有情形时，代价函数近似可表示为 $(R(d, \theta_1)-1)^2+(R(d, \theta_2)-1)^2=1$，已知 $P[\theta_1]=P[\theta_2]=0.5$，此时贝叶斯代价是多少?

7.9 令 $X:N(\mu, 1)$，其中 $\mu=\mu_1=1/2$ 或 $\mu=\mu_2=-1/2$。令 $H_1: \mu=-1/2$ 和 $H_2: \mu=1/2$，定义两种操作

a_1: H_1为真(H_2为假), a_2: H_2为真(H_1为假)。X 的样本空间为 $\Omega=\{-\infty, \infty\}$, 令 $S_1=\{-\infty, 0\}$ 及 $S_2=\{0, \infty\}$, 考虑两种互斥事件 $E_1=\{X\in S_1\}$ 和 $E_2=\{X\in S_2\}$。

(a)计算四种概率 $P(E_i|\mu_j)$ $i=1, 2; j=1, 2$。

(b)确定四种可能判决函数 d_i, $i=1, \cdots, 4$

(c)假定代价函数 $l(a_1, \mu_1)=0$, $l(a_1, \mu_2)=2$, $l(a_2, \mu_1)=5$, $l(a_2, \mu_2)=0$, 试计算(b)中各种判决函数的风险。哪个判决函数是不可接受的, 即至少有一个其他判决函数控制(好于)它?

7.10 有两个正态集合 X_1: $N(\mu_1, \sigma^2)$及 X_2: $N(\mu_2, \sigma^2)$, 在显著性水平 5% 时检验 H_1: $\mu_1=\mu_2$ 对 H_2: $\mu_1\neq\mu_2$, 试描述这个检验。

7.11 有两个正态集合 X_1:$N(\mu_1, \sigma^2)$及 X_2: $N(\mu_2, \sigma^2)$, 在显著性水平 5% 时检验 H_1: $\mu_1=\mu_2$ 对 H_2: $\mu_1>\mu_2$, 试描述这个检验。

7.12 调整 H_1: $\mu_1=\mu_2$ 对 H_2: $\mu_1<\mu_2$ 后重做习题 7.11。

7.13 假定有 n 个独立同分布正态随机变量观测 $X_1, \cdots, X_n$, 我们希望基于 n 个样本检验 H_1: $\sigma_X^2=\sigma_0^2$ 对 H_2: $\sigma_X^2\neq\sigma_0^2$, 描述此检验并证明最简单的检验可以用 χ^2 分布实现。

7.14 令 X: $N(\mu, \sigma^2)$, 其中已知 $\mu=1$ 或 0。基于单个观测判决假设 H_1: $\mu=1$ 对 H_2: $\mu=0$。

(a)证明似然比检验形式为: 如果 $\Lambda\triangleq\exp(\sigma^2(X-1))>k$, 则判决 H_1: $\mu=1$ 为真。

(b)证明等效检验为: 若 $X>c\triangleq k\sigma^2+0.5$, 则判决 H_1: $\mu=1$ 为真。

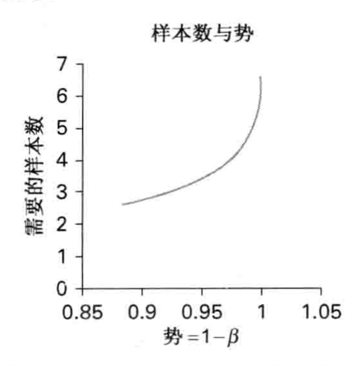
(c)给定 $\alpha=0.02$, 其中 $\alpha=P[H_1$ 为假$|$ H_1 为真$]$, 证明区分判决域的常量 c 由 $c=z_{0.02}\sigma+1$ 给出, 其中 $z_{0.02}$ 为标准正态累积密度函数的第二百分点。

(d)令 $\sigma=1$, 证明 $c=-2.05+1=-1.05$。

7.15 令 X:$N(\mu,1)$表示均值已知为 $\mu=\mu_1=3$ 或 $\mu=\mu_2=1$ 的集合, 取 X 的 n 个独立同分布观测$\{X_i, i=1, \cdots, n\}$, 令 H_1: $\mu=\mu_1=3$ 及 H_2: $\mu=\mu_2=1$, 证明似然比检验为若满足 $\hat{\mu}>(2n)^{-1}\ln(k)+2\triangleq c_n$, 则 H_1 为真, 其中 $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i$, 常量 c_n 由显著性水平 α 确定。寻求用 μ_1, n 和 z_α 表示的 c_n 的通用表达式, z_α 为 $N(0, 1)$分布的 α 百分位数。假定 $n=10$, 当 $\alpha=0.01$ 时的 c_n 是多少?

7.16 (习题 7.15 的继续)在习题 7.15 中, 若 n 为未知, 同时已知 $\alpha=0.02$ 及 $\beta=0.01$ 时计算 n 值。

7.17 (习题 7.16 的继续)保持 α 为 $\alpha=0.02$, 证明达到给定势的样本数如下图所示[提示: 应用 Excel 中的 NORMINV(概率, 均值, 标准差)]。



7.18 (方差比较的 F 检验)F 检验在检验两个正态集合方差(或标准差)是否相同时非常有用。典型地, 我们检验假设 H_1:$\sigma_1=\sigma_2$ 对 H_2:$\sigma_1\neq\sigma_2$, 当输入为两个正态集合 $N(\mu_1, \sigma_1^2)$和 $N(\mu_2, \sigma_2^2)$时, F 检验可以在线比较样本方差之比。因此, 假定有 P_1 的 m 个样本$\{X_{1i}, i=1, \cdots, m\}$, P_2 的 n 个样本$\{X_{2j}, j=1, \cdots, n\}$, 因为要保持样本方差相互独立, 我们没有将样本混合。用多个程序中的一个来计算输入现实$\{x_{1i}, i=1, \cdots, m\}$和$\{x_{2j}, i=1, \cdots, n\}$的样本方差数值, 用符号 s_1^2 和 s_2^2 表示为 $s_1^2\triangleq(m-1)^{-1}\sum_{i=1}^{m}(x_{1i}-\bar{x}_1)^2$(自由度 DOF$=m-1$)及 $s_2^2\triangleq(n-1)^{-1}\sum_{i=1}^{n}(x_{2j}-\bar{x}_2)^2$(自由度 DOF$=n-1$), 其

中 $\bar{x}_1 = m^{-1}\sum_{i=1}^{m} x_{1i}$ 和 $\bar{x}_2 = n^{-1}\sum_{j=1}^{n} x_{2j}$ 为样本均值。我们指定显著性水平 α，算法如下：(1)计算 $F' = s_1^2/s_2^2$，(2)比较 F' 与 $F_{\alpha/2, v_1, v_2}$，其中 $F_{\alpha/2, v_1, v_2}$ 为显著性 α 时 $m-1$ 和 $n-1$ 自由度的 F 分布的临界值。当检验 $H_1: \sigma_1 = \sigma_2$ 对 $H_2: \sigma_1 > \sigma_2$ 时，若 $F' > F_{\alpha, v_1, v_2}$ 时，判决 H_1 为假。

当检验 $H_1: \sigma_1 = \sigma_2$ 对 $H_2: \sigma_1 < \sigma_2$ 时，若 $F' < F_{1-\alpha, v_1, v_2}$ 时，判决 H_1 为假。

当检验 $H_1: \sigma_1 = \sigma_2$ 对 $H_2: \sigma_1 \neq \sigma_2$ 时，若 $F < F_{1-\alpha/2, v_1, v_2}$ 或 $F > F_{\alpha/2, v_1, v_2}$ 时，判决 H_1 为假。

作为一个练习，首先生成两个相同方差 σ 高斯随机数集，然后生成两个不同方差 σ 高斯随机数集，用在线计算器，如 BioKin 统计计算器来检验 F 检验的有效性。

7.19 (推广 F 检验到多集合)另一个应用 F 检验的方式是检验不同集合统计上是否相似，我们将在本题及下题中讨论基于统计背景的此类检验。假定有 k 个集合，每个集合有 n_i 样本，有 $\sum_{i=1}^{k} n_i = n$，令第 i 个集合中第 j 个样本表示为 Y_{ij}，令集合样本均值定义为 $Z_i \triangleq n_i^{-1}\sum_{j=1}^{n_i} Y_{ij}$，$i = 1, \cdots, k$，且 $\{Y_{ij}, j = 1, \cdots, n_i; i = 1, \cdots, k\}$ 为 n 个独立随机变量。在各集合中，n_i 个随机变量为独立同分布，因此对于所有 j 个样本，有 $\mathrm{Var}[Y_{ij}] = \sigma_{Y_i}^2$。通过定义 Z_i，我们得到 k 个正态随机变量 $\{Z_i, i = 1, \cdots, k\}$，其方差为 $\sigma_{Z_i}^2$，$i = 1, \cdots, k$，且所有样本的均值为 $\hat{\mu}_Z \triangleq k^{-1}\sum_{i=1}^{k} Z_i$。证明当 $\sum_{i=1}^{k}\left(\dfrac{Z_i - \hat{\mu}_Z}{\sigma_{Z_i}}\right)^2$ 服从 χ_{k-1}^2。令 $\mathrm{Var}[Y_{ij}] = \sigma_{Y_i}^2$，$i = 1, \cdots, K$，证明 $\sigma_{Z_i}^2 = \sigma_{Y_i}^2/n_i$，$i = 1, \cdots, k$，解释为什么 $\sum_{i=1}^{k}(Z_i - \hat{\mu}_Z)^2$ 有时称为集合内变化率或集合间变化率。

7.20 (习题 7.19 的继续)证明 $\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\dfrac{Y_{ij} - Z_i}{\sigma_{Y_i}}\right)^2$ 服从 χ_{n-k}^2。在何种假设下，$\sum_{i=1}^{k}\left(\dfrac{Z_i - \hat{\mu}_Z}{\sigma_{Z_i}}\right)^2 \Big/ \sum_{i=1}^{k}\sum_{j=1}^{n_i}((Y_{ij} - Z_i)/\sigma_{Y_i})^2$ 可以写成 $\sum_{i=1}^{k} n_i(Z_i - \hat{\mu}_Z)^2 / \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - Z_i)^2 = \chi_{k-1}^2/\chi_{n-k}^2$？最后通过分子除以 $k-1$ 及分母除以 $n-k$ 可生成 F 统计量，即

$$
\begin{aligned}
F_{k-1, n-k} &= (n-k)\sum_{i=1}^{k} n_i(Z_i - \hat{\mu}_Z)^2/(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - Z_i)^2 \\
&= (n-k)\chi_{k-1}^2/(k-1)\chi_{n-k}^2
\end{aligned}
$$

7.21 (F 检验)从参考文献[7-6]中我们得到如下真实数据来检验吸烟者和非吸烟者不同水平的氧气吸收能力，有5 项：

类　型	吸收率均值	吸收率标准差	各类型的人数
非吸烟者在无烟环境(1)	3.17	0.74	200
非吸烟者在有烟环境(2)	2.72	0.71	200
轻度吸烟者(3)	2.63	0.73	200
中度吸烟者(4)	2.29	0.70	200
重度吸烟者(5)	2.19	0.72	200

假设 H_1 为 5 项中的空气流动没有区别；备选假设为至少有一项呼吸统计与其他有明显差别①。在显著性水平为 0.05 时计算是否接受或拒绝假设。

7.22 (开方检验)植物学家试图通过两种豌豆的杂交来检验孟德尔的遗传法则，根据孟德尔法则，后代中四分之三应为绿色(支配颜色)，四分之一应为黄色(隐性)，在 880 株植物中，植物学家观测到 639 绿色种子和 241 黄色种子。令 H_1：绿色基因②占支配地位，H_2：绿色基因不占支配地位，在显著性水平 0.05 时判决假设的真假。

7.23 (t 检验)给定两组来自相同方差的正态分布现实，应用 t 检验检测 $H_1: \mu_1 = \mu_2$ 对 $H_2: \mu_1 \neq \mu_2$。

组 1：

① 注意到若我们判决假设不成立，我们还是不知道哪一项对判决负有责任。

② 一种转移遗传特性的基因。

−5.980e −1 −9.290e −1 −8.340e −2 1.020e +0 6.780e −1 2.890e −1 1.430e −1 −2.060e +0 1.260e +0 1.670e +0

组 2：

6.270e −1 2.640e +0 1.530e +0 5.920e −1 1.910e +0 5.050e −1 7.660e −1 2.760e −1 3.070e +0 8.550e −1

7.24　证明皮尔逊拟合性检验中的统计量 V 在 H_1 时的期望为 $E[V|H_1]=l-1$。

7.25　证明皮尔逊拟合性检验中的统计量 V 在 H_2 时的期望为 $E[V|H_2]>l-1$。

7.26　考虑两方差同等性检验的 F 检验，绘出 $m=8$，$n=5$ 时的检验统计量对方差比图形，求解显著性水平为 0.05 时的判决域。

7.27　从正态集合 P_1 中取 m 个样本，从正态集合 P_2 中取 n 个样本，检验两个方差的同等性。证明当 H_1 为真时，Λ 可以写成

$$\Lambda = A(m,n)\frac{\left(\frac{(m-1)}{(n-1)}F_{m-1,n-1}\right)^{m/2}}{\left(1+\frac{(m-1)}{(n-1)}F_{m-1,n-1}\right)^{(m+n)/2}}$$

其中 $A(m,n)\triangleq(m+n)^{(m+n)/2}m^{-m/2}n^{-n/2}$。

7.28　（该题与习题 7.13 相同，故略去。）

7.29　尽管年长者声明在赌博游戏中硬币是“公正的”，但年轻人在多次损失后表示怀疑，他相信硬币是有偏的，即 P[正面] = 0.75。在 50 次投掷实验中有 35 次出现“正面”，15 次出现“背面”，在显著性水平 0.05 时检验假设 P[正面] = 0.75。

7.30　随机变量 X 的 24 个观测值按大小排序为 $Y_1<Y_2<\cdots<Y_{24}$，估计 30 百分点。

7.31　在 25 个样本中求解中值的 98% 置信区间。

7.32　在显著性水平 $\alpha=0.05$ 时检验基于下列数据游程：

P_1：$S_1=\{-0.32, 1.05, 0.77, 0.23, -0.66, -2.03, -0.82, 1.97, -0.32, 1.12\}$

P_2：$S_2=\{3.2, -10.5, -7.7, -2.3, 6.6, 20.3, 8.2, -19.7, 3.2, -11.2\}$

判决 S_1 和 S_2 是否来自同一集合？

7.33　考虑 H_1：集合 P_1 与集合 P_2 相同对 H_2：集合不相同，在显著性水平 $\alpha=0.05$ 时基于下列数据游程。

P_1：$S_1=\{-2, -6, 8, 4, 2, -4, 6, 2, -8, -2\}$

P_2：$S_2=\{3, 7, 1, -9, -3, 5, -7, 3, 9, -1\}$

S_1 和 S_2 是否来自同一集合？从你的结论是否可以说明，此游程是不适合的？

参考文献

7-1　A. M. Mood and F. A. Graybill, *Introduction to the Theory of Statistics*, 2nd edition. New York: McGraw-Hill, 1963.

7-2　A. Papoulis, *Probability &Statistics*. Englewood Cliffs, NJ: Prentice Hall, 1990.

7-3　R. Walpole, R. Myers , and K. Ye, *Probability and Statistics for Engineers and Scientists*, 7th edition. Delhi: Pearson Education, 2002.

7-4　A. L. Garcia, *Probability Statistics and Random Processes for Electrical Engineering*, 3rd edition. Upper Saddle River, NJ: Prentice Hall, 2008

7-5　(Online) BioKin at: http://www.biokin.com/tools/fcrit.html

7-6　S. A. Glanz, *A Primer on Biostatistics*, 3rd edition. McGraw-Hill, New York, 1992.

7-7　N. J. Solkind, *Statistics for People Who (Think They) Hate Statistics*, 2nd edition. Sage Publications, 2010.

7-8 G. W. Snedecor and W. G. Cochran, *Statistical Methods*, 8th edition. Ames, IA: Iowa State Press, 1989.

7-9 (Online) NISTI/SEMATECH *e-Handbook of Statistical Methods*, http://www. itl. nist. gov/div898/handbook/, date

7-10 (Online) "The T-Test," available at http:// www. socialresearchmethods. net/kb/stat t. php

7-11 D. W. Zimmerman, "A Note on Interpretation of the Paired-Sample t-Test," *Journal of Educational and Behavioral Statistics*, Vol. 22, No. 3, pp. 349-360, 1997.

7-12 Student (W. S. Gossett), "The Probable Error of the Mean," *Biometrica*, Vol. 6, No. 1, pp. 1-25, 1908.

7-13 S. J. Coakes and L. G. Steed, *SPSS: Analysis for Windows: Version* 7. 0, 7. 5, 8. 0 *for Windows*, Wiley, Brisbane, Australia, 1999.

7-14 H. B. Mann and D. R. Whitney, "On a Test Whether One of Two Random Variables is Stochastically Larger Than the Other," *Annals of Mathematical Statistics*, Vol. 18, pp. 50-60, 1947.

7-15 (Online) Student's t-distribution, available at http://en. wikipedia. org/wiki/student's、tdistribution

7-16 Chernoff and E. L. Lehmann, "The Use of maximum Likelihood Estimates in χ^2 tests for Goodness-of-fit," *The Annals of Mathematical Statistics*, Vol. 25, pp. 579-586, 1954.

7-17 R. L. Plackett, "Karl Pearson and the Chi-Square Test," *International Statistical Institute (ISI)*, Vol. 51, No. 1, pp. 59-72, 1983.

7-18 K. Pearson, "On the Criterion that a Given System of Deviations from the Probable. . . Have Arisen from Random Sampling," *Philosophical magazine*, Series 550, Vol. 302, pp. 157-175, 1900.

7-19 J. Neyman and E. S. Pearson, "On the Use and Interpolation of Certain Test Criteria for Purposes of Statistical Inference," *Biometrica*, Vol. 20, pp. 175-240, 1928.

7-20 J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions of the Royal Society of London*, Series A, Vol 231, pp. 289-337, 1933.

7-21 O. Zeitouni, J. Ziv, and N. Mershav, "When is the GLRT Optimal?" IEEE Transactions on Information Theory, Vol. 38, pp. 1597-1601, 1991.

7-22 F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics*, Vol 1, pp. 80-83, 1945.

7-23 S. S. Wilks, "Order Statistics," *Bulletin of the American Mathematical Society*, Vol 54, pp. 6-50, 1948.

7-24 S Siegel, *Non-parametric Statistics for Non-Statisticians: A Step-by-Step Approach*, Upper Saddle River, NJ: Wiley, 2009.

第8章 随机序列

随机序列用于对信号与图像处理、数字控制以及通信等领域中的采样数据进行建模，也用于对诸如经济学变量、计算机寄存器的内容、甚至简单如掷硬币(伯努利试验)、或者计算机网络的某条链路上的数据包等固有的离散数据进行建模。在上述每种情况中，随机序列都是从用户的角度对这些数据源不可预知的行为进行建模。本章将学习随机序列及其重要性质。我们将会看到，一个随机(统计)序列可以看成随机变量的无限维向量①。因此随机序列介于有限维随机向量(参见第5章)和连续时间随机函数之间。连续时间随机函数被称为随机过程，将在下一章讨论。

另一种推广随机向量的方法是将索引参数的数量扩展为两个，由此得到随机矩阵，是在图像处理领域非常有用的数学模型。当随机矩阵的规模增长到无限大，可以得到一个二维的随机序列，可以用于图像和地球物理学信号处理中许多理论问题的研究。虽然本书不讨论图像处理，但是随机序列的基本概念可以推广到二维情况。三维和四维随机序列对视频和其他时空信号的不可预知属性进行建模也非常有效。

8.1 基本概念

在介绍这部分的时候，需要回顾并且扩展第1章中介绍的关于概率公理的某些基础知识。这是因为现在必须经常性地同时处理无限数量的随机变量，即随机序列。这部分知识的学习将从随机序列的定义开始，然后举一些简单的例子。

定义8.1-1 设$(\Omega, \mathscr{F}, P)$为一概率空间。令$\zeta \in \Omega$。用$X[n, \zeta]$表示建立在序号集Z上的样本空间Ω到复值序列空间中的映射。如果对每个确定的整数$n \in Z$，$X[n, \zeta]$为一随机变量，则$X[n, \zeta]$是一个随(统计)序列。序号集Z是所有的整数，$-\infty < n < +\infty$，必要时以0填充。

图8.1-1显示了样本空间$\Omega = \{1, \cdots, 10\}$的情况。可见对于一个固定的实验结果$\zeta$，$X[n,\zeta]$是一个普通的数值序列，即离散变量$n$的确定(非随机)函数。通常称这种数值序列是随机序列的实现或一个样本序列，记为$X_\zeta[n]$，在不至于混淆的情况下也可简记为$x[n]$。图8.1-1中画出了10个样本序列，对应于每个实验结果$\zeta \in \Omega$。换个角度，如果n不变，而将ζ作为变量，则$X[n,\zeta]$是一个随机变量②。因此，所有这些现实$(-\infty < n < +\infty)$的集合以及概率空间构成随机序列。本文经常但并不总是，将随机序列记为$X[n]$。当有必要明确样本空间Ω中某个实验结果ζ时，仍然采用记号$X[n, \zeta]$。注意这里对时间变量n用方括号是沿用离散时间信号处理的惯例。

---

① 在现实世界中，所有序列都是有限的。但是如果现实世界的序列相比其内在相关性而言足够长，无限长的序列模型基本上是准确的，除非所研究的恰好是现实世界序列的开头或结尾部分。

② 概率论基础教材中将随机变量的独立同分布序列记为$X_n(\zeta)$。但是这里的随机序列允许变量之间的依赖性。

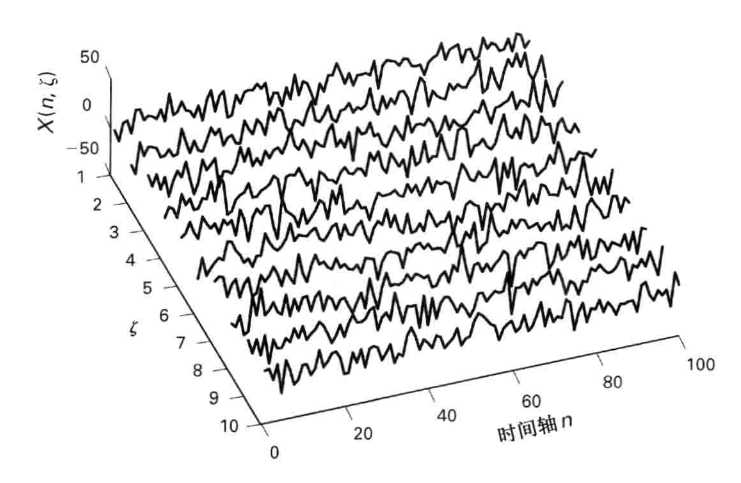


图 8.1-1 随机序列 $X(n,\zeta)$ 概念演示，其中 ζ 域(即采样空间 Ω)仅包含 10 个值

下面举一些随机序列的简单实例。

例 8.1-1 (可分离随机序列)令 $X[n,\zeta] \triangleq X(\zeta)f[n]$，其中 $X(\zeta)$ 是一个随机变量，$f[n]$ 是一个给定的确定(常规)序列。这种随机序列是可分离的随机变量(函数)和常规序列的乘积。也可以写为 $X[n]=Xf[n]$，按照随机变量的惯例省略实验结果 ζ。可以发现所有的样本序列都是其他样本按比例缩放的结果，而比例因子即为随机变量 X。

例 8.1-2 (随机幅度与相位的正弦波)令 $X[n,\zeta] \triangleq A(\zeta)\sin(\pi n/10+\Theta(\zeta))$，其中 A 和 Θ 为定义在概率空间 $(\Omega,\mathscr{F},P)$ 上的随机变量，也可以写为 $X[n]=A\sin(\pi n/10+\Theta)$。

这两个简单的随机序列是由确定性分量构成的，然而在其他方面它们也是"确定性"的。它们具有不寻常的属性，从概率论的角度，它们未来的取值由当前和过去的值完全确定。在例 8.1-1 中，一旦观察到 $X[n]$ 在某个确定的 $f[n]$ 时取值，例如 $n=0$，则由常规序列 $f[n]$ 是确定且已知的，因而整个随机序列 $X[n]$ 成为确定的。可见随机序列 $X[n]$ 在给定其 $n=0$ 处取值的情况下是条件已知的。例 8.1-2 的情况稍微复杂一点，但是用相同的方法足以表明，给定两个(非退化的)观测值，例如在 $n=0$ 和 $n=5$ 处的取值，就可以确定随机变量 A 和 Θ 的取值；序列 $X[n]$ 在 $n=0$ 和 $n=5$ 处的观测已知时成为条件已知或者完全可预知的。这种确定性的随机序列不能作为通信信道上噪声的有效模型，因为真正的噪声可不那么容易预知。

下面的例子将展示如何由随机向量构造一种更普遍、但是仍然"确定"的随机序列。

例 8.1-3 (有限支集的随机序列)设 $X[n,\zeta]$ 定义如下：

$$X[n,\zeta] \triangleq \begin{cases} X_n(\zeta), & 1 \leqslant n \leqslant N \\ 0, & \text{其他} \end{cases}$$

因为除了 $n \in [1,N]$ 之外 $X[n]=0$，因此称 $X[n]$ 是有限支集。由于这种有限支集特性，可以将该随机序列建模为一个随机向量 $\boldsymbol{X}=[X_1,X_2,\cdots,X_N]^{\mathrm{T}}$，然后运用矩阵代数丰富的计算方法，例如协方差矩阵和线性变换，如同第 5 章所介绍的。许多随机序列可以采用这种方法模拟，当然必须考虑 $N\to\infty$ 时 $\boldsymbol{X}$ 的极限来对更一般的随机序列建模。

例 8.1-4 (随机序列的树状图)设随机序列 $X[n]$ 定义在 $n \geqslant 0$ 上，而且只能取 M 个离散

的值，即0，1，2，…，$M-1$。进一步假设初始值固定为$X[0]=0$。那么可以用树状图来表示该随机序列的样本序列的演变，在每个节点$n=0$，1，2，…的分支数为M，如图8.1-2所示。

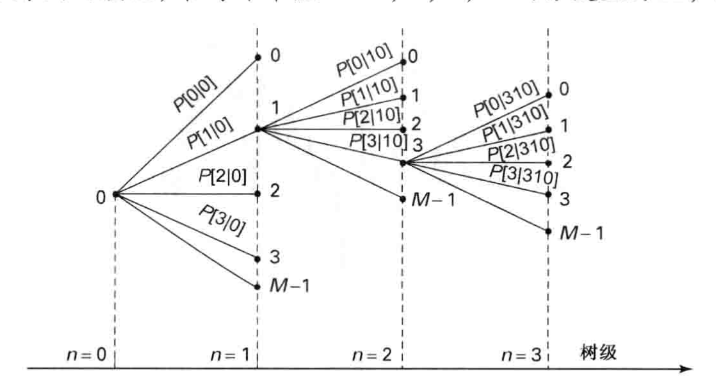


图8.1-2　离散幅度随机序列的树状图

在树的第n级，节点的值是可能的样本序列值$x[n]$，其分支标识为$i=0,\cdots,M-1$。样本序列由节点的取值序列标识，该取值序列是从根节点$n=0$开始沿着树的一条路径。如果将路径串$i_1i_2i_3\cdots$表示为M进制数值$0.i_1i_2i_3\cdots$，那么可以将这条路径所到达的点称为样本空间的结果[①] $\zeta\in[0,1]=\Omega$。最后可以用条件概率$P[X[n]=m_i|\{X[k]$对于$k\leqslant n-1$在相同的路径上$\}]$标识分支，在图8.1-2中记为$P[i_n|i_{n-1}i_{n-2}\cdots i_1 0]$。那么树的$n$级的任何节点值的概率可以由回溯到路径上根节点的所有分支概率的乘积确定。注意所有截止到n时刻一致的样本序列对应于样本空间$\Omega=[0,1]$中的一个半径为$\frac{1}{2}M^{-n}$的邻域。

这个例子显示了对于给出了构成随机序列的随机变量集合的概率分布信息的情况，如何构造一个一致的样本空间。注意因为所有随机变量都是两两独立的，即联合独立，而且概率分布不随时间变化，因此树上的分支标识都是不变的，实际上，树退化为一级。这种情况在概率论中被称为离散独立分布随机变量序列。将之稍加推广可以得到以下定义。

定义8.1-2　独立随机序列是指，对所有正整数N，在任何时刻n_1，n_2，…，n_N的随机变量都是联合独立的。

独立随机序列在我们的理论中扮演着重要的角色，因为它们相对易于分析，是构成更加复杂且精密模型的基础，而且利用计算机中的随机数发生器很容易得到一个近似的样本序列。此外，对连续时间数据采样得到的离散数据，如果样本之间的距离足够大，统计独立往往是非常有效的近似。

图8.1-3给出了实际噪声序列的一个片段，图8.1-4是其中一部分的特写，揭示了其离散时间本质和精细的“随机性”。这个片段也可以取自噪声序列的任何地方，而其统计性质相同。这种不一般的性质暗示了某种形式的“稳定性”，后面马上就要定义（参见定义8.1-5）。注意，产生此片段的连续随机变量看起来并不是独立的。实际上它们是相关的，因此必须采用一个N阶概率分布统计描述噪声序列的这个片段。按这种方法，将需要一个无穷阶的累积分布函数（CDF）来描述整个随机序列！

---

① 例如，令$M=8$，考虑八进制数值0.1200…0…。这表示$X[1]=1$，$X[2]=2$，之后所有的值都是0。

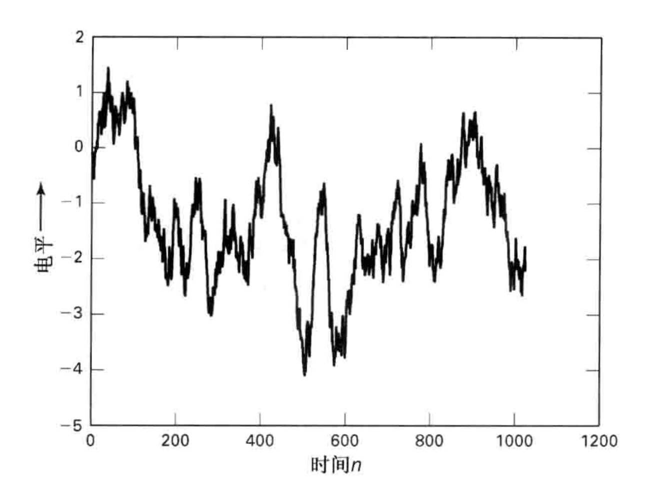


图8.1-3 随机序列的一个样本序列的例子

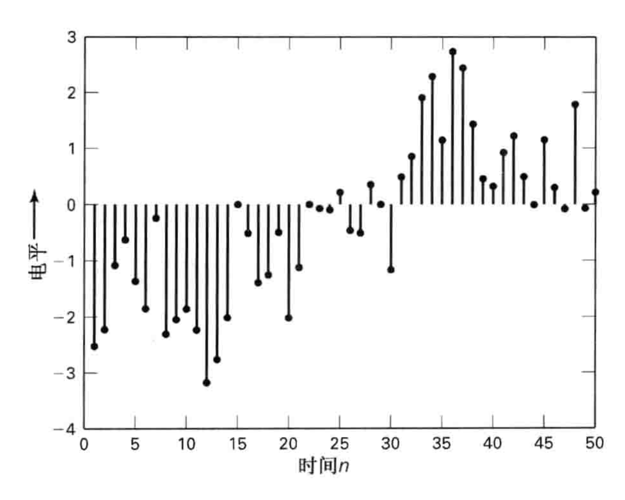


图8.1-4 部分样本序列的特写

为了处理无限长度的随机序列，必须能够计算事件的无限交集①的概率，例如事件$\{X[n]<5$，对所有正的$n\}$可以表示为$\cap_{n=1}^{+\infty}\{X[n]<5\}$，或者按照德·摩根定律，用无限并集表示为$(\cup_{n=1}^{\infty}\{X[n]\geqslant 5\})^c$。这要求能够定义并处理事件无限集合的概率，而这产生一个有关概率测度公理3的问题：也就是对于$AB=\phi$为空集

$$P[A\cup B]=P[A]+P[B]\quad(\text{公理}3) \tag{8.1-1}$$

通过迭代使用此公式可以将此结论推广到

$$P\left[\bigcup_{n=1}^{N}A_n\right]=\sum_{n=1}^{N}P[A_n]$$

对于任意有限正整数N，假设$A_iA_j=\phi$对所有$i\neq j$，这称为有限可加。它允许通过极限计算$\lim_{N\to\infty}P[\cup_{n=1}^{N}A_n]$，但上面需要的是$P[\cup_{n=1}^{\infty}A_n]$，其中$A_n\triangleq\{X[n]\geqslant 5\}$。对于一般函数，这两个量可能不一定相同，即$\lim_{N\to\infty}f(x_N)\neq f(\lim_{N\to\infty}x_N)$。为了使求极限可交换，需要给

① 请参阅1.4节回顾无限交集和并集的定义。此概念虽然简单，但是经常误解。

概率测度 P 引入某种连续性。这可以通过扩展或替换公理3，采用如下更普遍的无限(可数)可加公理。

公理4 (可数可加性)

$$P\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{+\infty} P[A_n] \tag{8.1-2}$$

对满足 $A_iA_j = \phi$，$i \neq j$ 的无限事件集合。

幸运的是在称为测度论[8-1](也可参考附录D)的数学分支中，证明了总是能够构造满足公理4的概率测度。俄罗斯数学家科尔莫戈罗夫[8-2]，经常被认为是现代概率论之父，已经证明了，如果定义了一种满足公理3的概率测度 P，即 P 满足有限加，总是可以扩展 P 使之满足可数可加公理4。下面看一个例子，之后再证明公理4等价于所要求的概率测度 P 的连续性。后面我们将假设概率测度满足公理4，称它们是可数可加的。

无限长伯努利试验

令 $\Omega = \{\mathrm{H}, \mathrm{T}\}$，即有两个实验结果变量 $\zeta = \mathrm{H}, \mathrm{T}$，$P[\mathrm{H}] = p$，$0 < p < 1$ 且 $P[\mathrm{T}] = q \triangleq 1 - p$。定义随机变量 W 为 $W(\mathrm{H}) \triangleq 1$ 和 $W(\mathrm{T}) \triangleq 0$，表示掷硬币中的胜利和失败。

令 Ω_n 为第 n 次掷硬币的样本空间(Ω 的第 n 个副本)，定义一个新的事件空间为无限叉积① $\boldsymbol{\Omega}_\infty \triangleq \times_{n=1}^{\infty}\Omega_n$。这对应于无限长掷硬币序列的样本空间，每次掷硬币的样本空间为 Ω_n。然后定义随机序列 $W[n, \boldsymbol{\zeta}] \triangleq W(\zeta_n)$，由此产生伯努利随机序列 $W[n]$，$n \geqslant 1$。这里结果变量 $\boldsymbol{\zeta}$ 由单次测试的结果给定，即 $\boldsymbol{\zeta} = [\zeta_1, \zeta_2, \cdots, \zeta_n, \cdots]$。

考虑无限维样本空间 $\boldsymbol{\Omega}_\infty$ 的概率测度。令 A_n 表示第 n 次试验的事件②，即 $A_n \in \mathscr{F}_n$，其中 $\mathscr{F}_n$ 是第 n 次试验的概率空间(Ω_n, $\mathscr{F}_n$, P)中的事件域，我们要求 $\cap_{n=1}^{\infty} A_n$ 为 $\mathscr{F}_\infty$ 中的一个事件，$\mathscr{F}_\infty$ 为 $\boldsymbol{\Omega}_\infty$ 中事件的 σ 域。为了使此事件域完备，必须为其增加所有此类事件的可数交集和并集。例如，要计算以下事件的概率

$$\{W[1] = 1, W[2] = 0\} \cup \{W[1] = 0, W[2] = 1\}$$

可以理解为两个形式为 $\cap_{n=1}^{\infty} A_n$ 的事件的并集，即 $\{W[1] = 1, W[2] = 0\} = \cap_{n=1}^{\infty} A_n$，按照 $A_1 = \{W[1] = 1\}$，$A_2 = \{W[2] = 0\}$ 及 $A_n = \Omega_n$，$n \geqslant 3$。则 $\mathscr{F}_\infty$ 必须包含所有此类事件以满足完备性。为了构造 $\boldsymbol{\Omega}_\infty$ 上的概率测度，从形如 $\boldsymbol{A}_\infty = \cap_{n=1}^{\infty} A_n$ 的集合开始，定义独立试验的情况

$$\boldsymbol{P}_\infty[\boldsymbol{A}_\infty] \triangleq \prod_{n=1}^{\infty} P[A_n]$$

然后利用公理4，以及 $\mathscr{F}_\infty$ 的每个成员可以表示为形如 $\cap_{n=1}^{\infty} A_n$ 的事件的可数并集和交集，将这种概率测度扩展到所有 $\mathscr{F}_\infty$。这样理论上构造了对应于无限长伯努利试验的概率空间($\boldsymbol{\Omega}_\infty$, $\mathscr{F}_\infty$, $\boldsymbol{P}_\infty$)，以及相应的伯努利随机序列

$$W[n, \boldsymbol{\zeta}] = W(\zeta_n), \qquad n \geqslant 1$$

---

① 这里的无限叉积 $\times_{n=1}^{\infty}\Omega_n$ 只是表示 $\boldsymbol{\Omega}_\infty$ 中的点包含所有无限长事件序列，每个序列在相应于某个 n 的 Ω_n 中。因此如果输出 $\boldsymbol{\zeta} \in \boldsymbol{\Omega}_\infty$，那么 $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \zeta_3, \cdots)$，其中对每个 $n \geqslant 1$，输出 $\zeta_n \in \Omega_n$(有限长伯努利试验的情况在1.9节中涉及)。

② 很可能只是一个单元素事件，即在这个二元情况中，只有一个输出。

已经看到如何为(无限长)伯努利随机序列构造样本空间 $\boldsymbol{\Omega}_\infty$，其中实验结果 $\boldsymbol{\zeta}$ 只是无限长的“H”和“T”序列。这个 $W[n]$ 是我们第一个非平凡的随机序列的例子。但是将每个随机变量 $W[n,\boldsymbol{\zeta}]$ 当成构成样本空间 $\boldsymbol{\Omega}_\infty$ 元素的无限维结果向量的函数，看起来可能有点勉强。就好像把问题故意复杂化，毕竟 $W[n,\boldsymbol{\zeta}]$ 只是 $W(\zeta_n)$。为了表明这种符号的复杂化是不可避免的，可以看看针对相关噪声通常采用的模型

$$X[n]=\sum_{m=1}^{n}\alpha^{n-m}W[m],\qquad n\geqslant 1 \tag{8.1-3}$$

其中 $W[n]$ 是刚刚生成的伯努利随机序列。为每个实验结果 $\boldsymbol{\zeta}$ 记录过滤后的输出 $X[n]$

$$X[n,\boldsymbol{\zeta}]=\sum_{m=1}^{n}\alpha^{n-m}W(\zeta_m)$$

可见每个 $X[n,\boldsymbol{\zeta}]$ 是与分量的数目持续增长(随 n)的函数，也就是说，$X[n]$ 的值依赖于实验结果 $\zeta_1,\zeta_2,\cdots,\zeta_n$。如果将处理确定数值 n 作为一个单独的问题，也就是一个独立的样本空间和概率测度，将会出现仍未回答的一致性问题。实际上需要应用科尔莫戈罗夫一致性定理，来证明我们的结论与具有(无限长)实验结果 $\boldsymbol{\zeta}$ 的单个样本空间 $\boldsymbol{\Omega}_\infty$ 一致①。

例 8.1-5 (相关噪声)考虑式(8.1-3)的随机序列，且 $|\alpha|<1$。取伯努利随机序列 $W[n]$ 作为输入，即 $W[n]=1$ 的概率为 p，$W[n]=0$ 的概率为 $q\triangleq 1-p$。希望求出每个大于 0 的 n 对应的 $X[n]$ 的均值。因为期望运算是线性的，可以得到

$$\begin{aligned}E\{X[n]\}&=E\left\{\sum_{m=1}^{n}\alpha^{n-m}W[m]\right\}\\&=\sum_{m=1}^{n}\alpha^{n-m}E\{W[m]\}\\&=\sum_{m=1}^{n}\alpha^{n-m}p=p\sum_{m=1}^{n}\alpha^{n-m}\\&=p\sum_{m'=0}^{n-1}\alpha^{m'}=p\frac{(1-\alpha^n)}{(1-\alpha)}\end{aligned}$$

由此产生的随机序列 $X[n]$ 不是独立随机变量的序列，通过计算其相关可以看出这一点

$$\begin{aligned}E\{X[2]X[1]\}&=E\{(\alpha W[1]+W[2])\,W[1]\}\\&=\alpha E\{W^2[1]\}+E\{W[2]\}E\{W[1]\}\\&=\alpha p+p^2\\&\neq(\alpha+1)p^2=E\{X[2]\}E\{X[1]\}\end{aligned}$$

随机变量 $X[2]$ 和 $X[1]$ 必定是关联的，它们甚至不是不相关的。

但是因为 $W[n]$ 是不相关的，所以容易计算其方差 $\{X[n]\}$ 为

---

① 这里对 $\boldsymbol{\Omega}_\infty$, $\boldsymbol{\zeta}$, $\boldsymbol{P}_\infty$ 用粗体有点浪费，只是为了避免混淆。显然 $\boldsymbol{\Omega}_\infty$ 不同于 $\lim\limits_{n\to\infty}\boldsymbol{\Omega}_n$。每个 Ω_n 的结果要么是 {H}，要么是 {T}，无论 n 变得多大。另一方面，$\boldsymbol{\zeta}\in\boldsymbol{\Omega}_\infty$ 的结果则是无限长的 H 和 T 的串。后面将不再使用粗体，即使 Ω 是由无限叉积产生而且其元素(结果)是无限长的串。

$$\begin{aligned}\operatorname{Var}\{X[n]\} &= \sum_{m=1}^{n} \operatorname{Var}\{\alpha^{n-m}W[m]\} \\ &= \sum_{m=1}^{n} \alpha^{2(n-m)}\operatorname{Var}\{W[m]\} \\ &= \frac{(1-\alpha^{2n})}{(1-\alpha^2)}pq\end{aligned}$$

此随机序列的动态属性可以用差分方程建模。由 $X[n-1]=\sum_{m=1}^{n-1}\alpha^{n-1-m}W[m]$，得到 $X[n]=\alpha X[n-1]+W[n]$，此结果明确显示 $X[n]$ 依赖于其邻点 $X[n-1]$①。因此相关噪声 $X[n]$ 可以由独立序列 $W[n]$ 用图 8.1-5 中的滤波器生成。由式(8.1-3)可以看出对于大的 n，$X[n]$ 是大量独立随机变量的和。因此根据中心极限定理，当 $n\to\infty$ 时将趋近于高斯分布，其均值为 $p\dfrac{1-\alpha^n}{1-\alpha}$，方差为 $pq\dfrac{1-\alpha^{2n}}{1-\alpha^2}$。

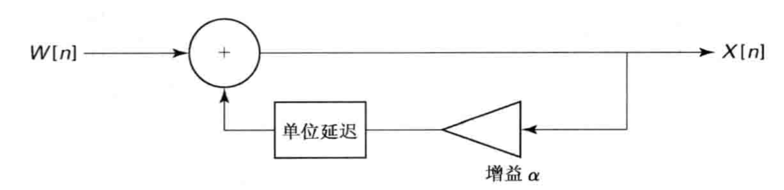


图 8.1-5　由非相关序列 $W[n]$ 生成相关噪声 $X[n]$ 的反馈滤波器

零均值相关高斯噪声可以用相同的模型生成。用 $W[1]$，$W[2]$，…，$W[n]$，…表示零均值独立同分布的高斯随机变量，且服从 $N(0,\sigma_W^2)$，则随机序列 $X[n]=\sum_{m=1}^{n}\alpha^{n-m}W[m]$ 是零均值高斯序列，其方差为

$$\operatorname{Var}\{X[n]\} = \frac{1-\alpha^{2n}}{1-\alpha^2}\sigma_W^2$$

其中 $\sigma_W^2=\operatorname{Var}\{W[n]\}$。同样地，这里由滤波器产生的序列是相关的，因为 $E\{X[2]X[1]\}=\alpha E\{W^2[1]\}=\alpha\sigma_W^2\neq E\{X[2]\}E\{X[1]\}=0$。

下面的例子介绍用 MATLAB 构造伯努利随机序列，然后将得到的样本序列输入一个一阶滤波器产生一个(更真实的)相关随机序列的样本。

例 8.1-6　(样本序列构造)采用 MATLAB 构造 $W[n]$ 一个样本序列。其 MATLAB 程序为

```
u = rand(40,1);
w = 0.5 >= u;
stem (w),
```

首先使用自带的函数“rand”产生一个 40 个元素的均匀分布的随机变量序列。第二行代码设置向量元素 $w[n]$，若 $u[n]\geq 0.5$，则设为 1，若 $u[n]<0.5$，则设为 0。因此 $w[n]$ 是 $p=0.5$ 的伯努利随机样本序列。相应的 MATLAB 的绘图如图 8.1-6 所示。

为了对 $X[n]$ 样本序列建模，将其记为 $x[n]$，可以对序列 $w[n]$ 用以下滤波器进行滤波：

① 这种公式中明确的依赖关系有时被称为直接依赖。

$$x[n] = \alpha x[n-1] + w[n]$$

该滤波器具有冲激响应 $h[n]=\alpha^n u[n]$，实现式(8.1-3)的线性运算。相应的 MATLAB 代码段为

```
b = 1.0;
a = [1.0 -alpha];
x = filter(b,a,w);
stem (x)
```

这里对 $\alpha=0.95$ 且具有400个元素的向量进行了计算。图8.1-7显示了前面40个值的初始瞬态值。图8.1-8显示了从 $n=350$ 开始50个点的近似稳态样本。

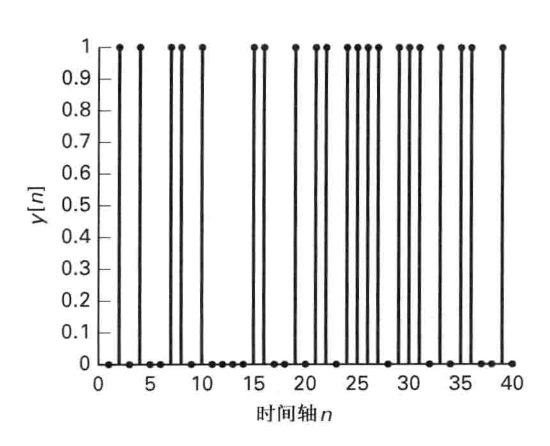


图8.1-6 伯努利随机序列 $W[n]$ 的一个样本序列 $w[n]$

图8.1-7 显示初始瞬态值的前40个点

注意随机序列 $X[n]$ 具有典型的类似噪声的特性。滤波器使构成 $X[n]$ 的随机变量相关，从而使样本序列 $x[n]$ 看起来更加“连续”。这个简单的例子称为自回归(AR)模型，在信号处理中被广泛用于对噪声和信号建模。注意前面例子中的确定性缺陷已经不存在了。这是因为伯努利输入序列为每个样本提供了新的独立值，确保了下一个样本不能由过去的预测得到。

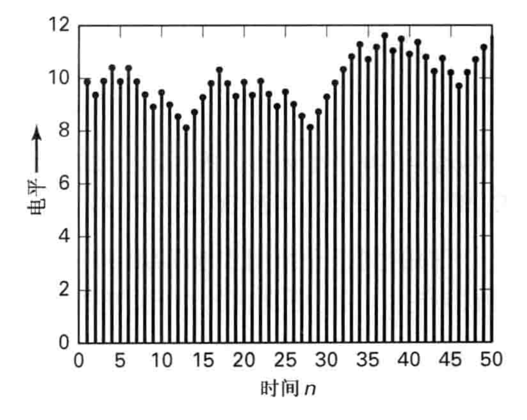


图8.1-8 从 $n=350$ 开始的50个点的片段

概率测度的连续性

在处理无穷数量的事件时已经发现概率测度的连续性非常有用。幸运的是，所期待的连续性是公理4具有可数可加性扩展的[参见式(8.1-2)]直接结果。

定理8.1-1 考虑事件 B_n 的递增序列，即对所有 $n\geq 1$，有 $B_n \subset B_{n+1}$，如图8.1-9所示。定义 $B_\infty \triangleq \cup_{n=1}^{\infty} B_n$，则 $\lim_{n\to\infty} P[B_n]=P[B_\infty]$。

证明 定义事件 A_n 的序列如下：

$$A_1 \triangleq B_1$$

$$A_n \triangleq B_n B_{n-1}^c, \qquad n > 1$$

则 A_n 不相交且对所有 N 有 $\cup_{n=1}^{N} A_n = \cup_{n=1}^{N} B_n$。同样地，因为 B_n 是递增的，所以 $B_N = \cup_{n=1}^{N} B_n$。因此

$$P[B_N] = P\left[\bigcup_{n=1}^{N} B_n\right] = P\left[\bigcup_{n=1}^{N} A_n\right] = \sum_{n=1}^{N} P[A_n]$$

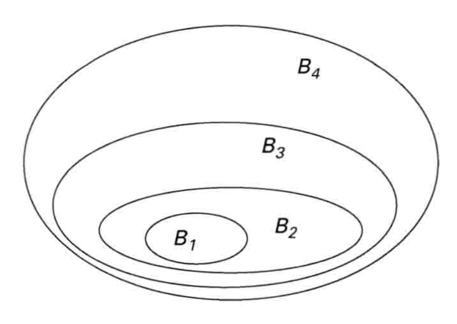


图 8.1-9 递增事件序列示例

以及

$$\lim_{n\to\infty} P[B_N] = \lim_{n\to\infty} \sum_{n=1}^{N} P[A_n]$$

$$= \sum_{n=1}^{+\infty} P[A_n] \quad \text{根据极限和的定义}$$

$$= P\left[\bigcup_{n=1}^{\infty} A_n\right] \quad \text{根据公理 4}$$

$$= P[B_\infty] \quad \text{根据 } A_n \text{ 的定义}$$

最后一步是根据 $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \triangleq B_\infty$。

推论 8.1-1 令 B_n 是事件的递减序列，即对所有 $n \geqslant 1$，有 $B_n \supset B_{n+1}$。则

$$\lim_{n\to\infty} P[B_n] = P[B_\infty]$$

其中

$$B_\infty \triangleq \bigcap_{n=1}^{\infty} B_n$$

证明 与定理 8.1-1 的证明类似，留给学生完成。

例 8.1-7 令 $B_n \triangleq \{X[k] < 2,\ 0 < k < n\}$，$n = 0, 1, 2, \cdots$。通俗地讲，$B_n$ 是指定范围 k 内 $X[k] < 2$ 的事件。显然 B_{n+1} 是 B_n 的子集，也就是对所有 $n = 0, 1, 2, \cdots$，有 $B_{n+1} \subset B_n$。同样地，如果令 $B_\infty \triangleq \{X[k] < 2,\ k \geqslant 0\}$，则 $B_\infty = \cap_{n=1}^{\infty} B_n$。因此按照上面的推论有

$$P[B_\infty] = \lim_{n\to\infty} P[B_n]$$

$$= \lim_{n\to\infty} P[X[0] < 2, \cdots, X[n] < 2]$$

因此，上述推论提供了一种计算包含无限多随机变量的事件方法，即通过求包含有限个随机变量的概率极限。这种求极限的方法在工程分析领域经常采用，而且通常不做清晰的解释(即不关注前面所说的一致性问题)。本节已经看到这种方法的正确性依赖于概率论的一个基本公理，即公理 4(可数可加性)。

接下来采用概率测度 P 的连续性证明关于累积分布函数的一个基本性质。

例 8.1-8 (右连续)累积分布函数是右连续的；也就是说，对于 $F_X(x) = P[X(\zeta) \leqslant x]$(参见 2.3 节 F_X 的性质 iii)，有

$$\lim_{n\to\infty} F_X\left(x + \frac{1}{n}\right) = F_X(x)$$

为了证明这一点，定义

$$B_n \triangleq \left\{\zeta: X(\zeta) \leqslant x + \frac{1}{n}\right\}$$

注意到 B_n 是递减事件序列，其中 $B_\infty \triangleq \cap_{n=1}^{\infty} B_n = \{\zeta: X(\zeta) \leqslant x\}$ 且

$$F_X\left(x + \frac{1}{n}\right) = P[B_n]$$

应用推论 8.1-1 可以得到

$$\begin{aligned}\lim_{n\to\infty} F_X\left(x + \frac{1}{n}\right) &= \lim_{n\to\infty} P[B_n] = P[B_\infty] \\ &= F_X(x)\end{aligned}$$

随机序列的统计特性

一个随机序列 $X[n]$ 称为统计确定的，如果对所有的 $N \geqslant 1$，以及 n，$n+1$，…，$n+N-1$ 时刻，其 N 阶累积分布函数是已知的，也就是已知

$$\begin{aligned}&F_X(x_n, x_{n+1}, x_{n+2}, \cdots, x_{n+N-1}; n, n+1, \cdots, n+N-1) \\ &\triangleq P[X[n] \leqslant x_n, X[n+1] \leqslant x_{n+1}, \cdots, X[n+N-1] \leqslant x_{n+N-1}]\end{aligned} \tag{8.1-4}$$

其中分号后面的变量 n，$n+1$，…，$n+N-1$ 表示此联合累积分布函数中 N 个随机变量的位置。注意对每个阶数 N，这是累积分布函数的无穷集合，因为必须知道在所有时刻 n，$-\infty < n < +\infty$ 的联合累积分布函数。为了使表达更加清晰，通常将联合累积分布函数简单记为

$$F_X(x_n, x_{n+1}, \cdots, x_{n+N-1}), \qquad \text{对于所有 } n, \qquad N \geqslant 1 \tag{8.1-5}$$

也可以定义非连续时间参数的 N 阶累积分布函数为

$$F_X(x_{n_1}, x_{n_2}, \cdots, x_{n_N}; n_1, n_2, \cdots, n_N)$$

这种统计特性看起来与整个随机序列的完整描述存在一定差距，因为在有限阶累积分布函数的无穷集合中没有一个是随机序列的完整描述。然而如果确定了所有有限时刻的有限阶联合分布，利用前面已经证明过的概率测度的连续性，可以通过对有限阶累积分布函数的极限运算求出包含无限个随机变量事件的概率。当然必须保证所选取的 N 阶累积分布函数本身是一致的。有时候可能非常简单，例如，组成随机序列的所有随机变量两两独立，比如说一个伯努利随机序列。

例 8.1-9 （一致性）所谓一致性，是指低阶的累积分布函数必须与高阶累积分布函数吻合。例如仅考虑 N 等于 2 和 3 的情况，必须满足

$$F_X(x_n, x_{n+2}; n, n+2) = F_X(x_n, \infty, x_{n+2}; n, n+1, n+2)$$

对于所有 n 以及所有 x_n 和 x_{n+2} 的值都成立。同样的 $N=1$ 累积分布函数也必须与 $N=2$ 一致。进一步地，一致性需要扩展到更高阶 N 的情况。

一致性可以通过结构模型来保证，如同例 8.1-6 中的伯努利随机序列滤波后的情况。如果对一个未知源的 N 阶累积分布函数的可疑集合，事实上要验证其一致性将是一件令人望而生畏的工作。因此在随机序列和随机过程中，构造模型扮演了非常重要的角色。

总之，目前已经看到了两种描述随机序列的方法：统计特征[参见式(8.1-4)]和基于随机

函数 $X[n,\zeta]$ 的直接描述。这里"统计"是指可以得到先验信息，至少理论上，可以通过估计 N 阶 $(N=1,2,3,\cdots)$ 累积分布函数来得到，也就是利用统计学。

如果 F_X 可导，则对于每个整数时刻 n 和正整数(阶)N，其 N 阶概率密度函数(pdf)可以通过以下方法计算：

$$f_X(x_n,x_{n+1},\cdots,x_{n+N-1};n,n+1,\cdots,n+N-1) \\ =\frac{\partial^N F_X(x_n,x_{n+1},\cdots,x_{n+N-1};n,n+1,\cdots,n+N-1)}{\partial x_n\partial x_{n+1}\cdots\partial x_{n+N-1}} \tag{8.1-6}$$

当仅考虑一个随机序列，有时省略下标 X。同样地，有时也省略时间记号，把 $f_X(x_n,x_{n+1},\cdots,x_{n+N-1};n,n+1,\cdots,n+N-1)$ 简写为 $f_X(x_n,x_{n+1},\cdots,x_{n+N-1})$。有时候需要处理复随机变量和随机序列。由此，可以用实随机变量的有序对 $X=(X_R,X_I)$ 来表示，通常写为 $X=X_R+jX_I$，其累积分布函数为

$$F_X(x_R,x_I)\triangleq P[X_R\leqslant x_R,X_I\leqslant x_I]$$

相应的概率密度函数为

$$f_X(x_R,x_I)=\frac{\partial^2 F_X(x_R,x_I)}{\partial x_R\partial x_I}$$

为简化表示，后面用 $f_X(x)$ 来表示 $f_X(x_R,x_I)$，若随机变量是复数，则各自的积分(对离散复数是求和)是 (x_R,x_I) 平面上的双重积分①。

随机序列的矩在许多应用中扮演重要角色。部分原因在于对于一大类随机序列(所谓的遍历序列，将在10.4 节介绍)，可以根据唯一的样本序列方便估计。一个随机序列的一阶矩或均值函数为

$$\mu_X[n]\triangleq E\{X[n]\}=\int_{-\infty}^{+\infty}xf_X(x;n)\mathrm{d}x \\ =\int_{-\infty}^{+\infty}x_nf_X(x_n)\mathrm{d}x_n$$

其中 $X[n]$ 为连续随机序列。在 n 时刻集合 $\{x_k,-\infty<k<+\infty\}$ 上的离散随机序列均值函数为

$$\mu_X[n]=E\{X[n]\}=\sum_{k=-\infty}^{+\infty}x_kP[X[n]=x_k] \tag{8.1-7}$$

对于混合随机序列的情况，与混合随机变量类似，可以方便地写为

$$\mu_X[n]=\int_{-\infty}^{+\infty}xf_X(x;n)\mathrm{d}x+\sum_{k=-\infty}^{+\infty}x_kP[X[n]=x_k] \tag{8.1-8}$$

实际上，利用斯蒂尔切斯积分概念[8-3]，两项可以按照累积分布函数 $F_X(x;n)$ 写成相同的形式 $\mu_X[n]=\int_{-\infty}^{+\infty}x\,\mathrm{d}F_X(x;n)$。

随机序列在 k 和 l 两个时刻取值的乘积 $X[k]X^*[l]$ 的期望值称为自相关函数，自相关函数是 k 和 l 的函数，其中 $-\infty<k,l<+\infty$，则

① 复随机序列被用做某些带通信号和噪声的等价基带模型。得到的复数模拟值可以用更低的采样速率来运行。

$$R_{XX}[k,l] \triangleq E\{X[k]X^*[l]\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_k x_l^* f_X(x_k,x_l;k,l)\mathrm{d}x_k\mathrm{d}x_l \tag{8.1-9}$$

当自相关函数存在时(通常存在，但是当然在某些情况积分可能不一定收敛)。我们大部分时候研究二阶随机序列，定义为平均功率 $E\{|X[n]|^2\}<\infty$ 有限，因此其相应的相关函数总是存在。后面将看到，在自相关函数定义中第二个参数的共轭可以极大简化复值随机序列。此外，定义中心化随机序列 $X_c[n]\triangleq X[n]-\mu_X[n]$，其均值为零，其自相关函数称为原始序列 $X[n]$ 的自协方差函数。定义为

$$K_{XX}[k,l] \triangleq E\{(X[k]-\mu_X[k])(X[l]-\mu_X[l])^*\} \tag{8.1-10}$$

直接从这些定义注意到，以下对称条件成立：

$$R_{XX}[k,l] = R_{XX}^*[l,k] \tag{8.1-11}$$

$$K_{XX}[k,l] = K_{XX}^*[l,k] \tag{8.1-12}$$

其称为埃尔米特对称。同时注意到

$$K_{XX}[k,l] = R_{XX}[k,l] - \mu_X[k]\mu_X^*[l] \tag{8.1-13}$$

方差函数定义为 $\sigma_X^2[n]\triangleq K_{XX}[n,n]$，表示 $X_c[n]$ 的平均功率。$X[n]$ 本身的功率已经在上面给出，等于 $R_{XX}[n,n]$。

例 8.1-10 (例 8.1-1 续)例 8.1-1 给出 $X[n]$ 的均值函数为

$$\mu_X[n] = E\{X[n]\} = E\{Xf[n]\} = \mu_X f[n]$$

其中 μ_X 是随机变量 X 的均值。其自相关函数为

$$R_{XX}[k,l] = E\{X[k]X^*[l]\} = E\{Xf[k]X^*f^*[l]\} = E\{|X|^2\}f[k]f^*[l]$$

因此自协方差函数为

$$\begin{aligned} K_{XX}[k,l] &= E\{|X|^2 f[k]f^*[l]\} - |\mu_X|^2 f[k]f^*[l] \\ &= E\{|X|^2 - |\mu_X|^2\}f[k]f^*[l] \\ &= E\{|X-\mu_X|^2\}f[k]f^*[l] \\ &= \sigma_X^2 f[k]f^*[l] \end{aligned}$$

其中 $\sigma_X^2=\mathrm{Var}(X)$。可见方差 $\sigma_X^2[n]$ 恰为 $\sigma_X^2|f[n]|^2$。

在下面的例子中将研究一个更符合随机性定义的序列。

例 8.1-11 (等待时间)考虑包含独立同分布随机变量 $\tau[n]$ $(n\geqslant 1)$ 的随机序列，每个随机变量满足式(2.4-16)给出的指数概率密度函数，即①

$$f_\tau(t;n) = f_\tau(t) = \lambda\exp(-\lambda t)u(t), \qquad n=1,2,\cdots$$

将时刻 n 之前的 $\tau[k]$ 和定义为

$$T[n] \triangleq \sum_{k=1}^{n}\tau[k] \tag{8.1-14}$$

---

① 注意 $\lambda=1/\mu$。

然后将 $T[n]$ $(n=1, 2, \cdots)$ 看成另一个随机序列。实际上，时间随机事件的到达通常以这种方式建模。称 $T[n]$ 为第 n 个事件到达的时间或等待时间，称 $\tau[n]$ 为间隔时间①，参见图 8.1-10。

在第 9 章将看到重要的泊松随机过程可以通过这种方式构造。这里首先需要基于式(8.1-14)的定义确定 n 时刻 $T[n]$ 的概率密度函数。因为 $\tau[k]$ 是独立的，可以应用式(4.7-3)得到 $T[n]$ 的概率分布函数为指数概率密度函数的 $(n-1)$ 重卷积。利用卷积求解 $T[2]$ 的概率密度函数，可以得到

$$f_T(t;2) = f_T(t) * f_T(t) = \lambda^2 t \exp(-\lambda t)u(t)$$

将此结果与指数概率密度函数再执行一次卷积，得到

$$f_T(t;3) = \frac{1}{2}\lambda^3 t^2 \exp(-\lambda t)u(t)$$

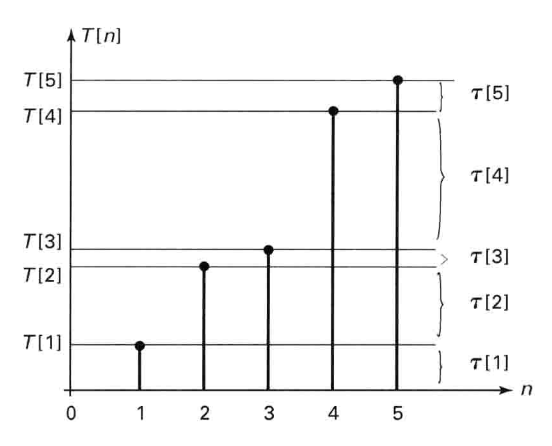


图 8.1-10 $T[n]$ 为到达时刻，$\tau[n]$ 为间隔时间

推广得到一般形式的 Erlang 概率分布函数

$$f_T(t;n) = \frac{(\lambda t)^{n-1}}{(n-1)!}\lambda \exp(-\lambda t)u(t) \tag{8.1-15}$$

Erlang 或 gamma 概率密度函数[8-4]广泛应用于通信网络中的等待时间问题，图 8.1-11 是 $n=3$ 和 $\lambda=1.0$ 时 MATLAB 绘制的曲线，对应于 $n=3$ 个到达事件的等待时间。

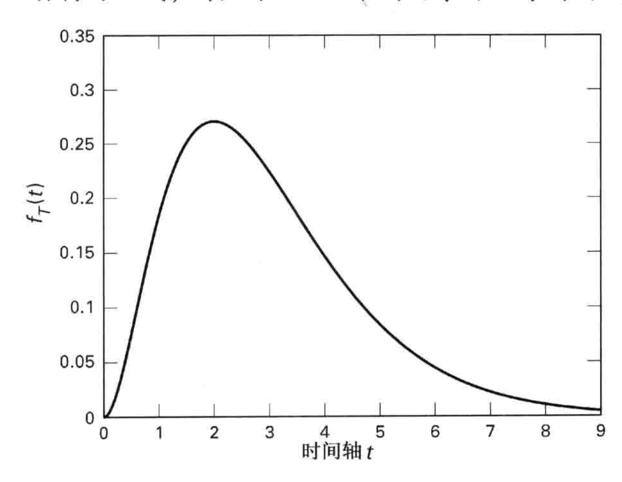


图 8.1-11 $\lambda=1$ 且 $n=3$ 时 Erlang 概率分布函数的曲线

可以利用数学归纳法证明此密度函数的正确性(参阅附录 A 的 A.4 节)。包含两个步骤：(1)首先证明 $n=1$ 时等式成立；(2)然后证明，如果等式在 $n-1$ 时成立，那么在 n 时也必然成立。结合这两个步骤，可以证明对于所有正整数 n，此结果都成立。

显而易见式(8.1-15)中 $f_T(t; 1)$ 是正确的，接下来假设在 $n-1$ 时式(8.1-15)是正确的。通过与指数概率密度函数卷积，以下验证在 n 时也是正确的：

---

① 请将 τ 看做大写，类似前面对随机变量及其取值的区分，即 $X=x$。

$$
\begin{aligned}
f_T(t;n) &= f_T(t;n-1) * \lambda\exp(-\lambda t)\,u(t) \\
&= \int_0^t \exp(-\lambda\tau)\frac{(\lambda\tau)^{n-2}}{(n-2)!}\lambda^2\exp(-\lambda(t-\tau))\mathrm{d}\tau\,u(t) \\
&= \lambda^n\exp(-\lambda t)\int_0^t \frac{\tau^{n-2}}{(n-2)!}\mathrm{d}\tau\,u(t) \\
&= \lambda^n\exp(-\lambda t)\frac{t^{n-1}}{(n-1)!}u(t)
\end{aligned}
$$

利用 $\tau[n]$ 独立同分布性质，可以计算其均值为

$$\mu_T[n] = n\mu_\tau = n(1/\lambda) = n/\lambda$$

重复利用式(4.3-18)的性质(A)计算和 $T[n]$ 的方差为

$$\mathrm{Var}[T[n]] = n\mathrm{Var}[\tau] = n/\lambda^2$$

下面介绍电子、通信和控制工程中应用最广泛的随机模型：高斯(正态)随机序列。它之所以被广泛应用，源于两个重要的事实：(1)中心极限定理(参见定理4.7-2)假定许多实际过程是近似高斯过程；(2)该模型对于检测、估计、滤波和控制理论中问题求解的数学处理非常简单。

定义8.1-3 一个随机序列 $X[n]$ 称为高斯随机序列，当其 N 阶($N\geqslant1$)累积分布函数(概率密度函数)是联合高斯分布。

注意到均值函数和协方差函数确定一个高斯随机序列同理，均值向量和协方差矩阵确定一个高斯随机向量(参见5.5节)。这是因为每个 N 阶分布函数恰为一高斯随机向量的累积分布函数，此高斯随机向量的均值向量和协方差矩阵可以用高斯随机序列的均值函数和协方差函数表示。

例8.1-12 (配对平均数)令 $W[n]$ 为一实值高斯独立同分布序列，对所有 n 其均值 $\mu_W[n]=0$，其自相关函数 $R_W[k,l]=\sigma^2\delta[k-l]$，$\sigma>0$，其中 δ 是离散时间冲激函数

$$\delta[n] \triangleq \begin{cases} 1, & n=0 \\ 0, & n\neq 0 \end{cases}$$

如果要构造一个协方差矩阵，那么对应一个任意 N 个分立样本的向量，其协方差矩阵将是对角矩阵。所以按照高斯分布特性，每个 N 阶概率密度函数可以分解为 N 个一阶概率密度函数的乘积。因此该随机序列的元素是联合独立的，或者可以称为独立(高斯)随机序列(参见定义8.1-2)。下面通过对 $W[n]$ 的当前与之前的值求和来构造随机序列 $X[n]$，即

$$X[n] \triangleq W[n] + W[n-1], \qquad -\infty < n < +\infty$$

这里 $X[n]$ 的所有 N 阶分布(因为根据定理5.6-1，高斯随机向量的线性变换也为一个高斯向量)也是高斯的；因此 $X[n]$ 也是一个高斯随机序列。可以非常方便地计算 $X[n]$ 的均值为

$$
\begin{aligned}
\mu_X[n] &= E\{X[n]\} = E\{W[n]\} + E\{W[n-1]\} \\
&= 0
\end{aligned}
$$

相关函数为

$$
\begin{aligned}
R_{XX}[k,l] &= E\{X[k]X[l]\} \\
&= E\{(W[k]+W[k-1])(W[l]+W[l-1])^*\} \\
&= E\{W[k]W[l]\} + E\{W[k]W[l-1]\}
\end{aligned}
$$

$$+E\{W[k-1]W[l]\}+E\{W[k-1]W[l-1]\}$$
$$=R_{WW}[k,l]+R_{WW}[k,l-1]+R_{WW}[k-1,l]+R_{WW}[k-1,l-1]$$
$$=\sigma^2(\delta[k-l]+\delta[k-l+1]+\delta[k-l-1]+\delta[k-l])$$

可以在$(k,\ l)$平面上绘出此自相关函数如图 8.1-12 所示，可以发现随机序列$X[n]$相关性的时间扩展。

从这个图形还可以看出，在对角线$l=k$上的自相关函数值为$2\sigma^2$，在对角线$l=k\pm1$上的自相关函数值为σ^2。由图 8.1-12 还可以明确$X[n]$不是一个独立随机序列。但是此协方差函数的带状属性强调了相关性在时间上局限于$(k-l)=\pm1$。除此以外都是不相关的，又因为是高斯的情况，所以是独立的随机变量。

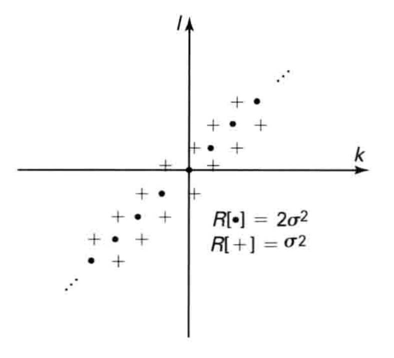


图 8.1-12 例 8.1-12 中三对角自相关函数的图形

例 8.1-13 （随机游动序列）继续讨论无限长伯努利试验，现在定义随机序列$X[n]$为在n次试验中成功（人头）次数的总和减去失败（字）次数的总和，取步进值为s，即

$$X[n]=\sum_{k=1}^{n}W[k] \qquad 且 \qquad X[0]=0$$

其中定义$W[k]=+s$对应于$\zeta=\mathrm{H}$，而$W[k]=-s$对应于$\zeta=\mathrm{T}$。

由此得到的序列为整数上以$X[0]=0$为出发点的随机游动模型。在每个相邻的时间单元，向左或者向右移动长度为s的一步。n次重复之后将处于对应于某个整数r的位置rs。如图 8.1-13 所示。

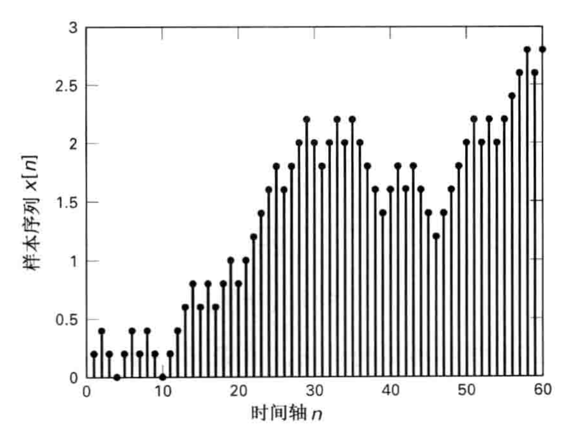


图 8.1-13 步进值$s=0.2$的随机游动$X[n]$的一个样本序列$x[n]$

如果有k次成功，显然有$(n-k)$次失败，则得到以下关系式：

$$rs=ks-(n-k)s$$
$$=(2k-n)s$$

这表明$k=(n+r)/2$，且r的取值使该式的右边为整数。由于$P[成功]=P[失败]=\frac{1}{2}$，所以有

$$P\{X[n]=rs\}=P[(n+r)/2\ 成功]$$
$$=\begin{cases}\binom{n}{(n+r)/2}2^{-n}, & (n+r)/2\ 为整数，r\leqslant n\\ 0, & 其他\end{cases}$$

由于$X[n]=W[1]+W[2]+\cdots+W[n]$，而且$W$是联合独立的，可以计算随机游动的均值和方差如下：

$$E\{X[n]\}=\sum_{k=1}^{n}E\{W[k]\}=\sum_{k=1}^{n}0=0$$

以及

$$\begin{aligned}E\{X^2[n]\}&=\sum_{k=1}^{n}E[W^2[k]]\\&=\sum_{k=1}^{n}0.5[(+s)^2+(-s)^2]\\&=ns^2\end{aligned}$$

如果对$X[n]$除以$\sqrt{n}$进行归一化，即定义

$$\tilde{X}[n]\triangleq\frac{1}{\sqrt{n}}X[n]$$

则根据中心极限定理4.7-2，可以得到$\tilde{X}[n]$的累积分布函数收敛于高斯(正态)分布$N(0, s^2)$。因此当n足够大，概率可以近似估计为

$$P[a<\tilde{X}[n]\leqslant b]=P[a\sqrt{n}<X[n]\leqslant b\sqrt{n}]\approx \mathrm{erf}(b/s)-\mathrm{erf}(a/s)$$

但是需要注意，当此概率较小时，要求n非常大以保持错误比例足够小，因为累积分布函数的小误差可能与概率值相当。实际上，这意味着正态近似不依赖于分布的尾部，而只依赖于中间部分，因此被称为中心极限定理。

同时还要注意，虽然对于任意n而言，$X[n]$不能看成近似高斯序列(例如，若n为偶数，则$X[n]$只能是偶数个s)，但是仍然可以近似地计算概率

$$\begin{aligned}P[(r-2)s<X[n]\leqslant rs]&=P\left[\frac{(r-2)s}{\sqrt{n}}<\tilde{X}[n]\leqslant\frac{rs}{\sqrt{n}}\right]\\&=\frac{1}{\sqrt{2\pi}}\int_{(r-2)/\sqrt{n}}^{r/\sqrt{n}}\exp(-0.5v^2)\mathrm{d}v\\&\approx 1/\sqrt{\pi(n/2)}\exp(-r^2/2n)\end{aligned}$$

其中r相对于$\sqrt{n}$足够小，1.11节有类似的结果。在最后一步的计算中，假设被积函数在区间$[(r-2)/\sqrt{n}, r/\sqrt{n}]$中近似为常量。

例8.1-11中的等待时间序列和例8.1-13中的随机游动序列都具有以下特点，即通过独立的分量以时间累积或通过增量构造得到。更一般地，可以定义独立增量性质。

定义8.1-4 如果对所有整数参数$n_1<n_2<\cdots<n_N$，对于任意$N>1$，增量序列$X[n_1]$，$X[n_2]-X[n_1]$，$X[n_3]-X[n_2]$，$\cdots$，$X[n_N]-X[n_{N-1}]$是两两独立的，则称随机序列具有独立增量性。

如果一随机序列具有独立增量性，则可以按照增量概率的乘积计算其N阶概率特性(概率分布函数和概率密度函数，参见习题8.10)。

与独立增量的渐变特性不同，许多随机序列具有恒定的统计特性，即不随时间或长度参数 n 变化，如果情况是这样的话，随机模型可以从两方面简化：首先模型是非时变的，其次通常少量的模型参数可以从已有的数据估计出来。

定义 8.1-5　若对所有的阶数 N 和变化参数 k($X[n]$, $X[n+1]$,…,$X[n+N-1]$)和($X[n+k]$, $X[n+k+1]$,…,$X[n+k+N-1]$)的联合累积分布函数是相同的，随机序列称为平稳的，即对于所有 $N\geqslant 1$

$$\begin{aligned}&F_X(x_n,x_{n+1},\cdots,x_{n+N-1};n,n+1,\cdots,n+N-1)\\&=F_X(x_n,x_{n+1},\cdots,x_{n+N-1};n+k,n+1+k,\cdots,n+N-1+k)\end{aligned}\tag{8.1-16}$$

对于所有的 $-\infty<k<+\infty$ 和所有 x_n 到 x_{n+N-1}。当概率密度函数以及对于离散幅度中的概率质量函数存在时，此定义同样成立。

回顾例 8.1-12，可以发 $X[n]$ 和 $W[n]$ 都是平稳随机序列。例 8.1-11 中的间隔时间 $\tau[n]$ 也是平稳的，但是随机到达或等待时间序列 $T[n]$ 显然不是平稳的，因为其均值和方差随时间 n 递增。

注意，平稳并不意味着样本序列看起来"相似"，甚或看起来是"噪声"①。而且，与数理统计中的平稳性概念不同，这里并不直接刻画随机序列实现的平稳性，只是用确定性函数来刻画，即累积分布函数、概率质量函数和概率密度函数等。

通常希望只基于随机序列的前两阶的信息，即均值和协方差函数来部分确定其特性。这已经在第 5 章对随机向量的处理中碰到过。在第 11 章讨论信号处理应用中的线性估计时还将碰到针对随机序列的情况。所以先定义一种弱化的平稳性，只包含均值和协方差(或者相关)函数。特别地，如果这两个函数都是平稳的，则称随机序列是广义平稳的(WSS)。

定义 8.1-6　在 $-\infty<n<+\infty$ 上定义的随机序列 $X[n]$ 称为广义平稳的(WSS)，如果：

(1) $X[n]$ 的均值函数对所有整数 n，$-\infty<n<+\infty$ 为常数，即

$$\mu_X[n]=\mu_X[0]$$

(2) 对所有时刻 k, l, $-\infty<k$, $l<+\infty$ 以及整数 $-\infty<n<+\infty$，协方差(相关)函数与偏移量 n 无关，即

$$K_{XX}[k,l]=K_{XX}[k+n,l+n]\tag{8.1-17}$$

称这种协方差(相关)函数具有平移不变性。若将[k, l]看成时间轴上两个样本的群或者集合，则在时间轴上上下移动此群，而协方差函数不发生改变。当均值函数为常量时，协方差和相关函数的平移不变性是等价的。否则不等价。对于常量均值函数，可以对协方差和相关函数验证性质(2)成立。

虽然所有的平稳序列都是 WSS，但是反过来则不成立。例如，即使一阶矩是常量，二阶矩是平移不变的，三阶矩完全有可能随时间变化，从而不满足平稳性。即随机序列是 WSS，但不

---

① 例如，假设通过掷一枚公平硬币一次进行伯努利试验，得到以下随机序列：若结果为"正面"则 $X[n]=1$ 对于所有 n，否则 $X[n]=W[n]$，即对于所有 n 为以平稳的白噪声。显然，样本序列看上去差别很大，但是此随机序列显然是平稳的。在第 10 章将讨论遍历性，通俗地说，就是可以由时间平均计算期望值(整体平均值)。此时样本函数倾向于具有相同的属性，即观测者在主观上感觉样本来自相同的源。

是平稳的。为了进一步说明这一点，有时可以将平稳性看成严格平稳，以避免与较弱的广义平稳相混淆。

定理 8.1-2 所有平稳随机序列都是广义稳定的。

证明 首先验证平稳随机序列的均值是常量。令 n 为任意值

$$\mu_X[n] = E\{X[n]\} = \int_{-\infty}^{+\infty} x f_X(x;n)\mathrm{d}x = \int_{-\infty}^{+\infty} x\, f_X(x;0)\mathrm{d}x = \mu_X[0]$$

其中因为 $f_X(x;\ n)$ 不依赖于 n。接下来验证协方差函数是平移不变的，为此先证明相关函数是平移不变的

$$\begin{aligned}R_{XX}[k,l] &= E\{X[k]X^*[l]\}\\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_k x_l^* f_X(x_k,x_l)\mathrm{d}x_k\mathrm{d}x_l\\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_{n+k}x_{n+l}^* f_X(x_{n+k},x_{n+l})\mathrm{d}x_{n+k}\mathrm{d}x_{n+l} \text{ ①}\\ &= R_{XX}[n+k,n+l]\end{aligned}$$

其中因为 $f_X(x_k,\ x_l)$ 不依赖于偏移 n，而且 x_i 均为哑变量。最后，利用式(8.1-13)和均值函数计算得到协方差函数也是平移不变的。因为对于任意 WSS 随机序列协方差函数都是平移不变的，可以为 WSS 序列定义唯一参数的协方差函数进行简化

$$\begin{aligned}K_{XX}[m] &\triangleq E\{X_c[k+m]X_c^*[k]\} = K_{XX}[k+m,k]\\ &= K_{XX}[m,0]\end{aligned} \tag{8.1-18}$$

同样地，可以简化相关函数。根据两个参数的相关函数，重写唯一参数的相关函数为

$$R_{XX}[m] = R_{XX}[k+m,k] = R_{XX}[m,0]$$

例 8.1-14 (WSS 协方差函数)例 8.1-12 的协方差函数是平移不变的，因此可以利用简化的公式记为 $K_{XX}[m]=\sigma^2(2\delta[m]+\delta[m-1]+\delta[m+1])$。

例 8.1-15 (有记忆二状态随机序列)按如下方法构造 $n\geqslant 0$ 处的二值(二元)随机序列 $X[n]$。对连续的每个 n(大于0)以及每个取值，递归地以某个给定的概率 $p(0<p<1)$ 设 $X[n]=X[n-1]$。否则以概率 $q\triangleq 1-p$ 设 $X[n]$ 为另一取值。令两种取值分别记为 a 和 b，且初始值 $X[0]=a$。当 $p=0.5$，则得到伯努利随机序列的特例。当 $p\neq 0.5$，则不是一个独立的随机序列，因为 $P_X(x_n|x_{n-1};\ n,\ n-1)\neq P_X(x_n;\ n)$，称随机序列是有记忆的。为了说明这一点，考虑 $p\approx 1.0$ 的情况，则 x_n 的值不同于 x_{n-1} 的条件转移概率 $P_X(x_n|x_{n-1};\ n,\ n-1)\approx 0$，但是无条件概率 $P_X(x_n;\ n)$ 并不受其限制。实际上，$P_X(e_n;\ n)$ 从期望值来看并不倾向任何取值，因为上面的转移规则并不区分取值。因此，至少称 $X[n-1]$ 为 $n-1$ 时刻的状态是合理的。实际上，生成此随机序列的规则可以概括为图 8.1-14 所示的状态转移图，其中有向弧上标识了从当前状态转移到下一状态的相对概率，这可以很容易的验证。这里将 p 称为保持概率。这是在 8.5 节将要研究的马尔可夫随机序列的第一个例子。

---

① 这里中间两行用到了简化的记号。它们一般不相等，因为 $f_X(x_k,x_l)$ 和 $f_X(x_{k+n},x_{l+n})$ 是在两个不同时间对的联合概率密度。用完整的记号 $f_X(x_k,x_l;k,l)$ 就可以表示得非常清楚。

下面的 MATLAB 源文件可以生成在 $n\geqslant1$ 时此类随机序列的样本函数：

```
function[w]=randmemseq(p,N,w0,a,b)
w=a*ones(1,N);
w(1)=w0;
for i=2:N
    rnum=rand;
    if rnum <p;
     w(i)=w(i-1);
     else
    if w(i-1)==a;
       w(i)=b;
    else
       w(i)=a;
    end
end
stem(1:N,w)
title('random sequence with memory')
xlabel('discrete time')
ylabel('level')
end
```

对应于取值 $b=1$，$a=0$ 以及不同取值的 p，相应的样本序列如图 8.1-15 至图 8.1-17 所示。注意当 p 接近 1 时，转移较少。对于接近 0.5 的 p 值，会呈现较多转移，显示较少的记忆。当 $p=0$ 时则每次都转移。

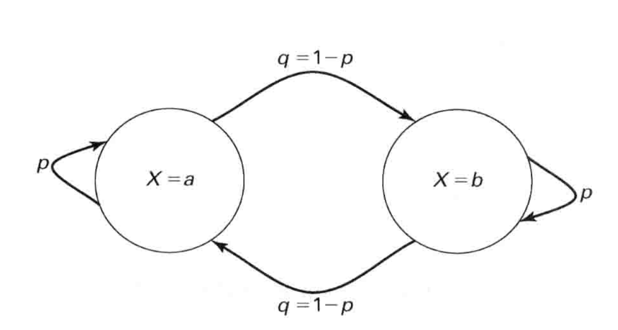


图 8.1-14　有记忆二值随机序列的状态转移图

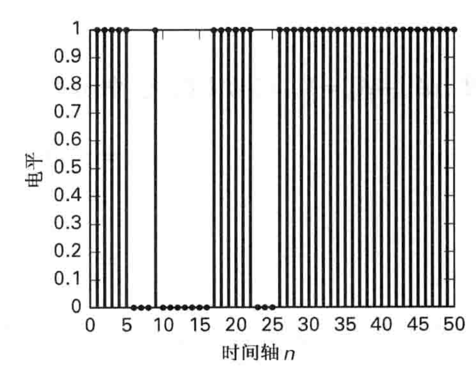


图 8.1-15　初始值 $X[1]=1$，保持概率 $p=0.8$

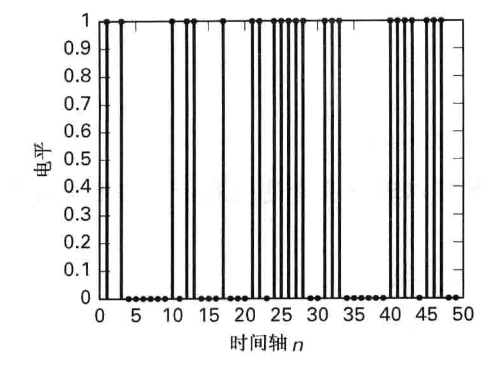


图 8.1-16　初始值 $X[1]=1$，保持概率 $p=0.5$，伯努利序列

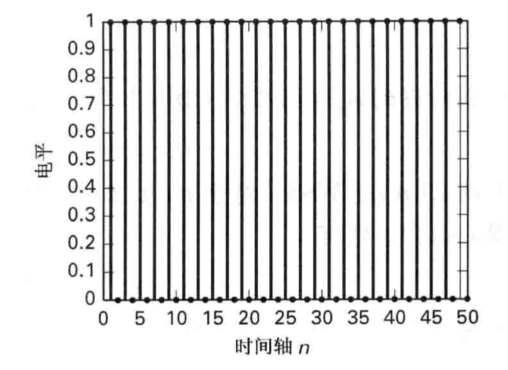


图 8.1-17　初始值 $X[1]=1$，保持概率 $p=0$

例 8.1-16　(有记忆随机序列的相关函数)假设上例中的有记忆随机序列已经持续了很长时间。下面将说明在这种情况下，将达到一个稳定状态，其中两种取值的概率随时间恒定且与初始值无关。这里假设稳定状态对所有有限时间成立。根据状态图所显示的对称性可见，必然有 $P_X(a)=P_X(b)=0.5$。现在进一步假设 $a=0$，b 仍然是原来的值，考虑两个离散时间 n 和 $n+k$ 的相关性如下：

$$\begin{aligned} R_{XX}[n,n+k] &= b^2 P_X(b,b;n,n+k) \\ &= b^2 P(X[n]=b)P(X[n+k]=b|X[n]=b) \\ &= (b^2/2)\,P(X[n+k]=b|X[n]=b) \end{aligned}$$

其中第一个等式成立的原因是所有包含 a 的项都是零，因为 $a=0$。现在 X 在时刻 n 和 $n+k$ 等于 b 的唯一可能是在这两个时间点之间发生偶数次转移，而此概率为

$$\begin{aligned} P\{\text{偶数次转移}\} &= \sum_{l=0,2,4,\cdots}^{k} \binom{k}{l} (1-p)^l p^{k-l} \\ &\triangleq A_e \end{aligned}$$

由此可见这相当于取"成功"="转移"、"失败"="不转移"的伯努利试验。因此交换伯努利试验中 p 和 q 的角色，只是对偶数次成功(转移)的概率进行相加。由此可以采用以下技巧近似估计 A_e 的值。定义

$$A_o \triangleq \sum_{l=1,3,5,\cdots}^{k} \binom{k}{l} (1-p)^l p^{k-l} (-1)^l$$

显然，$A_e - A_o = 1$，因为 l 在 A_o 中总是取奇数。同样地，注意到

$$\begin{aligned} A_e &= \sum_{l=0,2,4,\cdots}^{k} \binom{k}{l} (1-p)^l p^{k-l} (-1)^l \\ &= \sum_{l=0,2,4,\cdots}^{k} \binom{k}{l} (p-1)^l p^{k-l} \end{aligned}$$

其中第一个等式成立的原因是在 A_e 中 l 总是取偶数。现在可以看到根据伯努利定理，有

$$\begin{aligned} A_e + A_o &= \sum_{l=0}^{k} \binom{k}{l} (p-1)^l p^{k-l} \\ &= (2p-1)^k \end{aligned}$$

由此立刻得到 $A_e = (1/2)[(2p-1)^k + 1]$，因此

$$R_{XX}[n,n+k] = (b^2/4)\,[(2p-1)^k + 1]$$

这表明 $X[n]$ 是 WSS。对于 $p>1/2$ 的情况，该相关函数可以进一步清楚地表示。若定义 $\alpha \triangleq$ $\ln(2p-1)|$，则有

$$R_{XX}[k] = (b^2/4)\,[\exp(-\alpha\,|k|) + 1]$$

同时由于 $X[n]$ 的均值容易确定为 $b/2$，可以得到自协方差函数

$$K_{XX}[k] = (b^2/4)\,\exp(-\alpha\,|k|)$$

相应于三种不同取值的 p，显示此类序列的协方差函数的 MATLAB 源文件如下：

```
function[mc1,mc2,mc3]=markov(b,p1,p2,p3,N)
mc1=0*ones(1,N);
mc2=0*ones(1,N);
mc3=0*ones(1,N);
for i=1:N
   mc1(i)=0.25*(b^2)*(((2*p1-1)^(i-1)));
   mc2(i)=0.25*(b^2)*(((2*p2-1)^(i-1)));
   mc3(i)=0.25*(b^2)*(((2*p3-1)^(i-1)));
end
x=linspace(0,N-1,N);
plot(x,mc1,x,mc2,x,mc3)
title('covariance of Markov Sequences')
xlabel('Lag interval')
ylabel('covariance value')
```

对于 $p=0.8$, 0.5, 0.2 和 $b=2$ 的归一化协方差函数如图 8.1-18 所示。

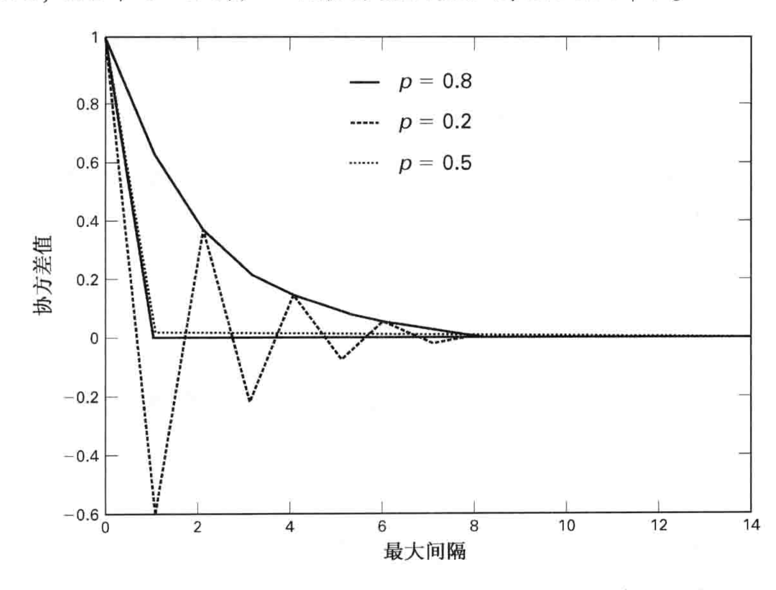


图 8.1-18　不同参数 p 取值的协方差函数(点之间用直线连接)

这种具有单步记忆的随机序列称为马尔可夫(1856 - 1922)随机序列。在 8.5 节将深入讨论这类随机序列。同时注意到例 8.1-5 所讨论的系统，即 $X[n]=\alpha X[n-1]+W[n]$，当 $W[n]$ 是一个独立随机序列时，也显示了单步记忆，因此也可以看成马尔可夫序列。

8.2 节简要回顾序列的线性系统理论，即离散时间线性系统理论。具有足够基础的读者可以跳过这一节。8.3 节将应用该理论研究随机序列输入线性系统的响应，它在通信、信号处理和控制系统方面应用广泛。

8.2　离散时间线性系统基础

本节介绍离散时间线性系统的一些基本概念。然后在下一节扩展到随机序列输入和输出的情况。这部分与连续时间线性系统理论非常相似，包括差分方程、傅里叶变换和拉普

拉斯变换等内容。在离散时间理论中的对应内容是差分方程、傅里叶变换(离散时间信号)和 Z 变换。

按照图 8.2-1，一个线性系统可以看成输入为无限长序列 $x[n]$，相应的输出为无限长序列 $y[n]$。以公式表示这种线性运算为

$$y[n]=L\{x[n]\} \tag{8.2-1}$$

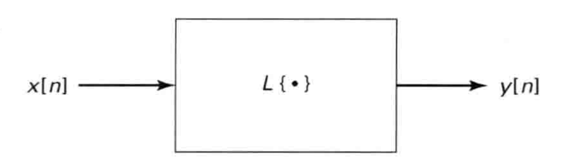


图 8.2-1 以 $x[n]$ 为输入，$y[n]$ 为输出的一般线性系统 $L\{\cdot\}$ 的系统框图

其中线性算子 L 满足后面针对离散时间信号情况的定义。这种表示看上出好像表明时刻 n 的 $x[n]$ 是唯一影响 n 时刻输出 $y[n]$ 的输入。实际上，所有输入值都可能对任意时刻 n 的输出具有潜在影响。这是为什么称 L 是算子①而不只是一个函数。下面的例子将明确这一点。数学家采用算子符号 $y=L\{x\}$ 来避免这种麻烦，但是却使得依赖于 n 为(时间)参数的 x 和 y 的函数比工程化表示更加容易混淆。

定义 8.2-1 称算子 L 的系统是线性的，当对所有可能的输入序列 $x_1[n]$ 和 $x_2[n]$ 以及所有可能的标量增益 a_1 和 a_2 都有

$$L\{a_1x_1[n]+a_2x_2[n]\}=a_1L\{x_1[n]\}+a_2L\{x_2[n]\}$$

通俗地说，线性系统对于加权和输入的响应为各个输出响应的加权和。线性系统的例子包括如下式表示的移动平均：

$$y[n]=0.33(x[n+1]+x[n]+x[n-1]),\qquad -\infty<n<+\infty$$

以及下式表示的自回归：

$$y[n]=ay[n-1]+by[n-2]+cx[n],\qquad 0\leqslant n<+\infty$$

其初始条件为0。这两个等式都是一般线性常系数差分方程(LCCDE)的特例

$$y[n]=\sum_{k=1}^{M}a_ky[n-k]+\sum_{k=0}^{N}b_kx[n-k] \tag{8.2-2}$$

例 8.2-1 (差分方程的解)考虑下面的二阶 LCCDE：

$$y[n]=1.7y[n-1]-0.72y[n-2]+u[n] \tag{8.2-3}$$

其中 $y[-1]=y[-2]=0$ 且 $u[n]$ 为单位阶跃函数。为了求 $n\geqslant 0$ 时此方程的解，首先求解齐次方程

$$y_h[n]=1.7y_h[n-1]-0.72y_h[n-2]$$

尝试 $y_h[n]=Ar^n$，其中 A 和 r 是确定性的②，由此得到

$$A(r^n-1.7r^{n-1}+0.72r^{n-2})=0$$

或

$$Ar^{n-2}(r^2-1.7r+0.72)=0$$

由此可见任何满足特征方程

$$r^2-1.7r+0.72=0$$

① 算子将函数(序列)映射到函数(序列)。

② 完整的线性差分方程求解问题的讨论可以参阅参考文献[8-5]。

的 r 为齐次方程的一般解。对本例而言有两个根，分别为 $r_1=0.8$ 和 $r_2=0.9$。通过线性叠加，一般齐次解形如①

$$y_h[n]=A_1r_1^n+A_2r_2^n$$

其中常数 A_1 和 A_2 可以由初始条件确定。

为了得到特解，首先观察到当 $n\geqslant 0$ 输入序列 $u[n]$ 等于 1。因此可以尝试特解为一个常数，即按照标准方法

$$y_p[n]=B,\qquad n\geqslant 0$$

得到

$$B-1.7B+0.72B=1$$

或者

$$B=1/(1-1.7+0.72)=1/(0.02)=50$$

更一般地，这种方法可以修改以针对任意在相邻时间间隔 $[n_1, n_2-1]$ 形如 $C\rho^n$ 的输入函数。只要假设解的相应形式并确定常数 C。在这种方法中，需要对每个时间间隔分别求解差分方程，然后通过把当前时间段最后的状态作为下一时间段的初始状态将各个时间间隔的解拼接起来。这里演示了起始于 $n=0$ 的时间间隔。完整的解为

$$\begin{aligned}y[n]&=y_h[n]+y_p[n]\\&=A_1(0.8)^n+A_2(0.9)^n+50,\qquad n\geqslant 0\end{aligned}$$

为了确定 A_1 和 A_2，首先估计式(8.2-3)在 $n=0$ 和 $n=1$ 处的值，利用 $y[-1]=y[-2]=0$ 结合初始条件得到 $y[0]=1$ 和 $y[1]=2.7$，由此得到线性方程

$$A_1+A_2+50=1\quad(n=0)$$

和

$$A_1(0.8)+A_2(0.9)+50=2.7\quad(n=1)$$

写成矩阵形式

$$\begin{bmatrix}1.0 & 1.0\\0.8 & 0.9\end{bmatrix}\begin{bmatrix}A_1\\A_2\end{bmatrix}=\begin{bmatrix}-49.0\\-47.3\end{bmatrix}$$

解得

$$\begin{bmatrix}A_1\\A_2\end{bmatrix}=\begin{bmatrix}32\\-81\end{bmatrix}$$

由此对 $n\geqslant 0$ 的完整解为

$$y[n]=32(0.8)^n-81(0.9)^n+50$$

如果系统在 $n<0$ 时是静止的，则所有时间的解为

$$y[n]=\{32(0.8)^n-81(0.9)^n+50\}\,u[n]$$

注意，上例中的 LCCDE 为一线性系统是因为初始条件，即 $y[-1]$，$y[-2]$ 为零，通常称为零初始条件。如果初始条件不为零，则 LCCDE 不是一个线性系统。更一般地，线性系统描述为如下时变冲激响应的叠加态：

$$h[n,k]\triangleq L\{\delta[n-k]\}$$

① 因为两个根在幅度上都小于 1，所以该解在时间轴 n 上是稳定的（参见参考文献[8-5]）。

通俗来讲，$h[n, k]$表示系统在 n 时刻对 k 时刻的冲激响应。只要将输入表示为 $x[n]=\sum x[k]\delta[n-k]$ 即可推导出结果，然后利用线性特性得出

$$\begin{aligned} y[n] &= L\left\{\sum_{k=-\infty}^{+\infty} x[k]\delta[n-k]\right\} \\ &= \sum_{k=-\infty}^{+\infty} x[k]L\{\delta[n-k]\} \\ &= \sum_{k=-\infty}^{+\infty} x[k]\, h[n,k] \end{aligned}$$

称为线性系统的叠加和表示。

许多线性系统由不变的单元构成，对任何时候到达系统的输入信号其响应是不变的。一个线性系统对延迟(移位)输入的响应也仅仅是延迟(移位)，则称为线性时不变(LTI)，或者等价的线性移不变(LSI)系统。更确切地有以下定义。

定义 8.2-2　线性系统 L 被称为移不变的，如对所有整数移位 k，$-\infty<k<+\infty$ 都有

$$y[n+k]=L\{x[n+k]\}\quad 对于所有 n \tag{8.2-4}$$

LSI 系统的一个重要性质是它们是由卷积[①]描述的，即 L 是一个卷积算子

$$y[n]=h[n]*x[n]=x[n]*h[n]$$

其中

$$h[n]*x[n]\triangleq\sum_{k=-\infty}^{+\infty}h[k]x[n-k] \tag{8.2-5}$$

同时

$$h[n]\triangleq L\{\delta[n]\}$$

称为冲激响应。相对于时变冲激响应 $h[n, k]$，若线性系统时不变，则 $h[n]=h[n, 0]$。

通俗地，可以说(仅对连续时间系统)如果知道一个 LSI 系统的冲激响应，那么通过卷积运算可以计算该系统对任意其他输入的响应。对于离散时间的情况，卷积运算是求和而不是积分，但是运算的其他方面是类似的。

虽然只要知道了系统的冲激响应，理论上可以确定任意输入的输出响应，实际上卷积运算可能烦琐且耗时。为了简化计算，而且促进理解，可以从频域来刻画 LSI 系统。首先定义序列的傅里叶变换(FT)如下所述。

定义 8.2-3　离散时间信号或者序列的傅里叶变换定义为以下无限和(若存在)：

$$X(\omega)=\mathrm{FT}\{x[n]\}\triangleq\sum_{n=-\infty}^{+\infty}x[n]\mathrm{e}^{-\mathrm{j}\omega n},\qquad -\pi\leqslant\omega\leqslant+\pi$$

$X(\omega)$是周期函数，周期为 2π。傅里叶变换的逆运算为

$$x[n]=\mathrm{IFT}\{X(\omega)\}=\frac{1}{2\pi}\int_{-\pi}^{+\pi}X(\omega)\mathrm{e}^{\mathrm{j}\omega n}\mathrm{d}\omega$$

---

① 在第 3 章计算两个独立随机向量和的概率密度函数的时候用过卷积运算。

可以看出，序列的傅里叶变换及其逆变换实际上是我们熟悉的傅里叶级数，只是序列x作为傅里叶系数，而傅里叶变换X充当了周期函数。因此，傅里叶级数的存在性和唯一性定理立刻可以用于这里的离散时间信号的傅里叶变换。注意频率变量ω有时称为归一化频率，因为如果序列$x[n]$是通过采样获得的，那么采样的周期已经丢失。这就好比采样周期为$T=1$，恰好与傅里叶变换$X(\omega)$的频率区间$[-\pi, +\pi]$一致①。

对于一个 LSI 系统，其傅里叶变换尤其重要，因为复指数是离散时间线性系统的本征函数，即

$$L\{\mathrm{e}^{\mathrm{j}\omega n}\}=H(\omega)\mathrm{e}^{\mathrm{j}\omega n} \tag{8.2-6}$$

只要冲激响应函数h是绝对可加的。对于 LSI 系统，这种绝对可加性很容易验证等价于有界输入有界输出(BIBO)稳定性[8-5]。

与连续时间系统定理类似，傅里叶变换的乘积对应于时间(空间)域的卷积。

定理 8.2-1 (卷积定理)卷积

$$y[n]=x[n]*h[n], \qquad -\infty<n<+\infty$$

在变换域上等价于

$$Y(\omega)=X(\omega)H(\omega), \qquad -\pi\leqslant\omega\leqslant+\pi$$

证明

$$\begin{aligned}
Y(\omega)&=\sum_{-\infty}^{+\infty}y[n]\mathrm{e}^{\mathrm{j}\omega n}=\sum_{-\infty}^{+\infty}(x[n]*h[n])\,\mathrm{e}^{-\mathrm{j}\omega n}\\
&=\sum_n\sum_k x[k]h[n-k]\mathrm{e}^{-\mathrm{j}\omega n}=\sum_n\sum_k x[k]h[n-k]\mathrm{e}^{-\mathrm{j}\omega(n-k+k)}\\
&=\sum_n\sum_k[x[k]\mathrm{e}^{-\mathrm{j}\omega k}h[n-k]\mathrm{e}^{-\mathrm{j}\omega(n-k)}]\\
&=\sum_k x[k]\mathrm{e}^{-\mathrm{j}\omega k}\left(\sum_n h[n-k]\mathrm{e}^{-\mathrm{j}\omega(n-k)}\right)\\
&=\sum_k x[k]\mathrm{e}^{-\mathrm{j}\omega k}H(\omega)\\
&=X(\omega)H(\omega)
\end{aligned}$$

因此离散时间线性移不变系统在频域上很容易理解，类似于连续时间 LSI 系统。与连续时间信号的拉普拉斯变换相对应，离散时间信号有Z变换。其定义如下所述。

定义 8.2-4 离散时间信号或序列的Z变换定义为无限和(若存在)

$$X(z)\triangleq\sum_{n=-\infty}^{+\infty}x[n]z^{-n} \tag{8.2-7}$$

其中z为此无限和的绝对收敛域中的复变量。

注意$X(z)$是复变量的函数，而$X(\omega)$是实变量的函数。两者的关系为$X(z)|_{z=\mathrm{e}^{\mathrm{j}\omega}}=X(\omega)$。由此可见，如果$Z$变换存在，则傅里叶变换只是将$Z$变换限制在复$z$平面的单位圆上

① 如果序列的采样周期为T，则(实际)角频率为$\boldsymbol{\Omega}=\omega/T$。

的特例。类似于定理8.2-1的证明，容易验证式(8.2-1)的卷积乘性质对于Z变换同样成立。与连续时间定理对应，一个LSI系统的冲激响应$h[n]$的Z变换$H[z]$称为系统函数。关于离散时间信号和系统更多信息，可以参阅参考文献[8-5]。

8.3 随机序列与线性系统

本节研究输入随机序列的线性系统。特别是线性系统和LSI系统对均值和协方差函数的变换。首先将考虑一般的非平稳随机序列的情况，然后针对特定的更常见的平稳序列。本节的主题可能是随机序列理论中应用最广泛的概念。具体应用包括：在通信中分析经过线性滤波器的信号与噪声、在数字信号处理中分析数字滤波器量化噪声，以及在控制论中发现输入扰动对非确定控制系统的影响等。

第一个主题是随机序列输入线性系统的意义。问题在于随机序列并不仅仅是单个序列，而是以参数ζ，采样空间上的一个点(输出)，标识的一簇序列。因此对每个固定的ζ，随机序列只是一个普通的序列，线性系统可能的输入之一。因此，当讨论一个输入随机序列的线性系统时，会很自然地说对样本空间Ω中的每一个点，输入相应的实现，即样本序列$x[n]$。然后考虑相应的输出$y[n]$作为与样本平面上的样本点ζ对应的样本序列①，从而联合地定义输出随机序列$Y[n]$。

定义8.3-1 对于随机序列$X[n]$和线性系统L，$Y[n]=L\{X[n]\}$表示对每个$\zeta\in\Omega$，有

$$Y[n,\zeta]=L\{X[n,\zeta]\}$$

等价地，对取自输入随机序列$X[n]$的每个样本函数$x[n]$，设$y[n]$为相应的输出随机序列$Y[n]$的样本序列，即$y[n]=L\{x[n]\}$。

这是处理随机输入系统最简单的方法。当输入样本序列的“不规范”会导致困难，即在这种情况下可能不一定能对每一个样本序列都定义系统操作。在“网上章节”的第10章将推广该定义并讨论系统操作的均方描述，它避开了这个问题，但必然更加抽象。

对于多数情况很难从线性系统输入的概率描述得到输出的概率分布。原因在于，冲激响应通常非常长(或者无限长)，所以要求用输入序列的高阶分布来确定输出的累积分布函数。换言之，若$Y[n]$依赖最近k个输入$X[n]$，…，$X[n-k+1]$，则必须知道X的k阶概率密度函数以计算Y的(甚至只是一阶)概率密度函数。矩函数的情况稍有不同。当系统为线性，输出随机序列的矩可以通过输入的等阶或低阶矩计算得到。部分缘于此，由输入的矩函数确定输出的矩函数显得非常有趣。对于实用并且重要的高斯随机序列的情况，已经知道完整的概率描述仅依赖于均值和协方差函数。实际上，由于线性系统可以对无限维向量构成的输入序列有效地执行线性变换，可以发现如果输入高斯序列，则输出序列的n阶分布同样遵守高斯定律。因此当输入高斯序列时，确定输出的一阶和二阶矩函数显得非常重要。

定理8.3-1 对于线性系统L和随机序列$X[n]$，输出随机序列$Y[n]$的均值为

$$E\{Y[n]\}=L\{E\{X[n]\}\} \tag{8.3-1}$$

只是两边都有定义。

---

① 对于给定ζ，$x[n]$，$y[n]$分别表示$X[n,\zeta]$，$Y[n,\zeta]$。

证明(正式) 因为 L 是线性算子，对每个样本序列，记输入输出对为

$$y[n]=\sum_{k=-\infty}^{+\infty}h[n,k]x[k]$$

或者

$$Y[n,\zeta]=\sum_{k=-\infty}^{+\infty}h[n,k]X[k,\zeta]$$

其中明确地表明了实验结果 ζ。如果对两边执行期望运算 E，得到

$$E\{Y[n]\}=E\left\{\sum_{k=-\infty}^{+\infty}h[n,k]X[k]\right\}$$

现在假设可以将算子 E 移到无限求和的里面，则得到

$$\begin{aligned}E\{Y[n]\}&=\sum_{k=-\infty}^{+\infty}h[n,k]E\{X[k]\}\\&=L\{E\{X[n]\}\}\end{aligned}$$

这可以写成

$$\mu_Y[n]=\sum_{k=-\infty}^{+\infty}h[n,k]\mu_X[k]$$

也就是输出的均值函数是线性系统对输入均值函数的响应。

这里有必要对期望和线性算子的交换进行说明。这并不总是可行的！例如，如果输入具有非零均值函数而线性系统是累加的，也就是

$$y[n]=\sum_{k=0}^{+\infty}x[n-k]$$

则均值的累加可能不收敛。那么这种交换就是无效的。在8.7节学习随机收敛时还会讨论这一点。到那时会看到一个LSI系统满足式(8.3-1)的充分条件是其冲激响应 $h[n]$ 是绝对可加的。

式(8.3-1)的一些特殊情况依赖于输入序列是否WSS以及系统是否为LSI。如果系统为LSI且输入至少是WSS的话，输出的均值为

$$E\{Y[n]\}=\sum_{k=-\infty}^{+\infty}h[n-k]\mu_X$$

现在由于 μ_X 是常数，可以将其从求和中移出来，得到

$$E\{Y[n]\}=\left[\sum_{k=-\infty}^{+\infty}h[k]\right]\mu_X \tag{8.3-2}$$

$$=H(z)|_{z=1}\,\mu_X \tag{8.3-3}$$

只要 $\sum_{k=-\infty}^{+\infty}|h[k]|$ 存在，也就是对任意BIBO稳定系统(参见8.2节)。

由此可见，在这种情况下，输出随机序列的均值是常数，等于LSI系统的直流增益或固定增益乘以输入序列的均值。

例 8.3-1 (低通滤波器)令具有以下系统函数的系统低通滤波器

$$H(z)=1/(1+az^{-1})$$

其中要求 $|a|<1$ 以使设想的因果滤波器满足稳定性(即收敛区域为 $|z|>|a|$, 包括单位圆)。那么若此滤波器的输入是一个 WSS 序列, 其输出的均值为

$$E\{Y[n]\}=H(z)|_{z=1}E\{X[n]\}$$
$$=(1+a)^{-1}\mu_X$$

下面计算算子为 L 的线性系统输出的协方差和相关函数:

$$Y[n]=L\{X[n]\}$$

引入输入和输出之间的互相关函数有助于问题的解决

$$R_{XY}[m,n]\triangleq E\{X[m]Y^*[n]\} \tag{8.3-4}$$

$$=E\{X[m]\,(L\{X[n]\})^*\} \tag{8.3-5}$$

现在, 为了分离出算子, 引入算子 L_n^*, 其冲激响应为 $h^*[n,k]$, 作用在时间参数 k 上, 将参数 n 看成常量。然后记 $R_{XY}[m,n]=E\{X[m]L_n^*[X^*[n]]\}=L_n^*E\{X[m]X^*[n]\}$。同样用 L_m 表示将 n 看成常量, 作用在时间参数 m 上的线性算子。算子 L_n^* 与线性代数中研究的共轭算子相关。

定理 8.3-2 令 $X[n]$ 和 $Y[n]$ 分别是线性算子 L_n 的输入和输出随机序列。令输入的相关函数为 $R_{XX}[m,n]$, 则互相关函数及输出的相关函数分别为

$$R_{XY}[m,n]=L_n^*\{R_{XX}[m,n]\}$$

和

$$R_{YY}[m,n]=L_m\{R_{XY}[m,n]\}$$

证明 记

$$X[m]Y^*[n]=X[m]L_n^*\{X^*[n]\}$$
$$=L_n^*\{X[m]X^*[n]\}$$

则

$$R_{XY}[m,n]=E\{X[m]Y^*[n]\}=E\{L_n^*\{X[m]X^*[n]\}\}$$
$$=L_n^*\{E\{X[m]X^*[n]\}\}$$
$$=L_n^*\{R_{XX}[m,n]\}$$

由此得出定理的第一部分。为了验证第二部分, 类似地将 $Y[m]$ 乘以 $Y^*[n]$ 得到

$$E\{Y[m]Y^*[n]\}=E\{L_m\{X[m]Y^*[n]\}\}$$
$$=L_m\{E\{X[m]Y^*[n]\}\}$$
$$=L_m\{R_{XY}[m,n]\}$$

将定理 8.3-2 的两部分结合, 可以得到用输入的相关函数表示的输出相关函数的算子表达式

$$R_{YY}[m,n]=L_m\{L_n^*\{R_{XX}[m,n]\}\} \tag{8.3-6}$$

将其代入时变冲激响应为 $h[n,k]$ 的系统的叠加求和公式, 有

$$R_{YY}[m,n]=\sum_{k=-\infty}^{+\infty}h[m,k]\left(\sum_{l=-\infty}^{+\infty}h^*[n,l]R_{XX}[k,l]\right) \tag{8.3-7}$$

这里$R_{XY}[m,n]$的叠加求和表示为

$$\begin{aligned}R_{XY}[m,n]&=L_n^*\{R_{XX}[m,n]\}\\&=\sum_{l=-\infty}^{+\infty}h^*[n,l]R_{XX}[m,l]\end{aligned}$$

而$R_{YX}[m,n]$的叠加求和表示为

$$R_{YX}[m,n]=\sum_{k=-\infty}^{+\infty}h[m,k]R_{XX}[k,n]$$

为了求相应的协方差函数的结果，注意到由于系统的线性以及式(8.3-1)，中心化输出序列是由于中心化输入序列产生的输出。将定理8.3-2应用于零均值序列，则对于协方差函数，立刻得到

$$K_{XY}[m,n]=L_n^*\{K_{XX}[m,n]\} \tag{8.3-8}$$

$$K_{YY}[m,n]=L_m\{K_{XY}[m,n]\} \tag{8.3-9}$$

和

$$K_{YY}[m,n]=L_m\{L_n^*\{K_{XX}[m,n]\}\} \tag{8.3-10}$$

由此得到下面的叠加求和

$$K_{YY}[m,n]=\sum_{k=-\infty}^{+\infty}h[m,k]\left(\sum_{l=-\infty}^{+\infty}h^*[n,l]K_{XX}[k,l]\right) \tag{8.3-11}$$

例8.3-2　(边沿检测器)令$Y[n]\triangleq X[n]-X[n-1]=L\{X[n]\}$为一阶(反向)差分算子，参见图8.3-1。此线性算子可用于定位某些随机数据中的脉冲噪声峰值。输出的均值为$E[Y[n]]=L\{E[X[n]]\}=\mu_X[n]-\mu_X[n-1]$。互相关函数为

$$\begin{aligned}R_{XY}[m,n]&=L_n\{R_{XX}[m,n]\}\\&=R_{XX}[m,n]-R_{XX}[m,n-1]\end{aligned}$$

输出的自相关函数为

$$\begin{aligned}R_{YY}[m,n]&=L_m\{R_{XY}[m,n]\}\\&=R_{XY}[m,n]-R_{XY}[m-1,n]\\&=R_{XX}[m,n]-R_{XX}[m-1,n]-R_{XX}[m,n-1]+R_{XX}[m-1,n-1]\end{aligned}$$

如果输入随机序列为WSS，且其自相关函数为

$$R_{XX}[m,n]=a^{|m-n|},\qquad 0<a<1$$

则上面的例子可以具体为

$$\mu_Y[n]=0$$

$$R_{XY}[m,n]=a^{|m-n|}-a^{|m-n+1|}$$

以及

$$R_{YY}[m,n]=2a^{|m-n|}-a^{|m-1-n|}-a^{|m-n+1|}$$

可见其仅仅依赖于 $m-n$。因此输出随机序列为 WSS，可记为(令 $k=m-n$)

$$R_{YY}[k]=2a^{|k|}-a^{|k-1|}-a^{|k+1|}$$

对于如图 8.3-2 给出的输入自相关函数中 $a=0.7$ 的情况，输出的自相关函数如图 8.3-3 所示。注意，边沿检测器倾向于对输入序列解相关。

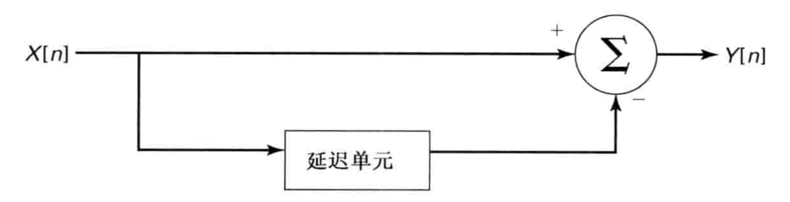


图 8.3-1 边沿检测器，当 $X[n]\approx X[n-1]$ 时输出趋近于 0，当 $|X[n]-X[n-1]|$ 较大时输出也较大

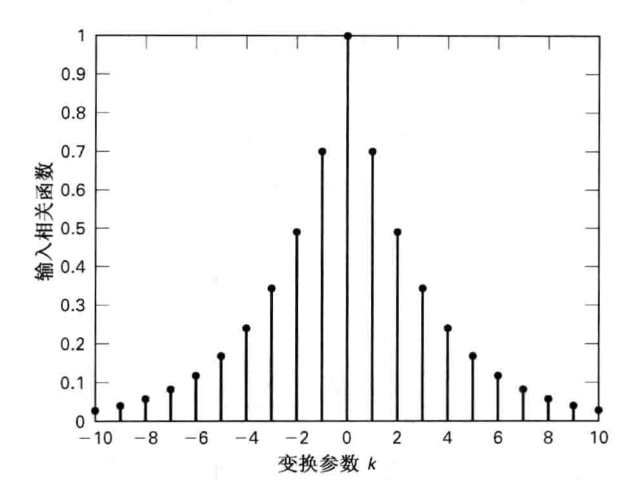


图 8.3-2 边沿检测器输入的相关函数，$a=0.7$

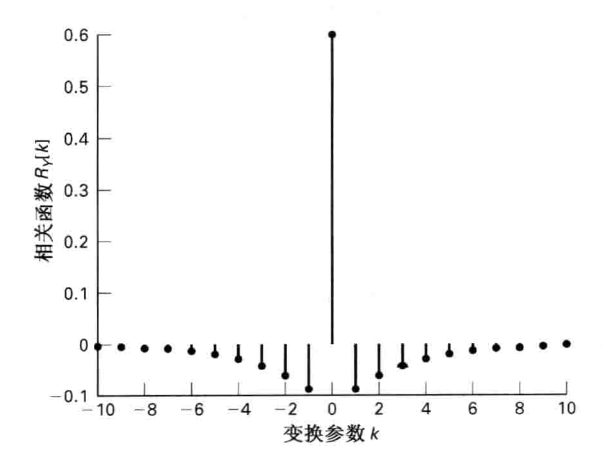


图 8.3-3 反向差分的相关函数 $R_{YY}[k]$(图中取 $a=0.7$)

例 8.3-3 (递归系统的协方差函数)取 $|\alpha|<1$，令

$$Y[n]=\alpha Y[n-1]+(1-\alpha)W[n] \tag{8.3-12}$$

其中 $n\geqslant 0$ 且 $Y[-1]=0$。因为初始条件为零，$n\geqslant 0$ 时系统等价为 LSI，因此可以通过卷积表

示L，其中

$$h[n]=(1-\alpha)\alpha^n u[n]$$

这里的$h[n]$是确定性一阶差分方程对应的冲激响应，即$h[n]$是以下方程的解

$$h[n]=\alpha h[n-1]+(1-\alpha)\delta[n]$$

其中$\delta[n]$是离散时间冲激序列。此解很容易通过递归求解或利用Z变换[①]得到。代入到式(8.3-1)，得到

$$\mu_Y[n]=\sum_{k=0}^{\infty}(1-\alpha)\alpha^k\mu_W[n-k],\quad 其中\ \mu_W[n]=0,\quad n<0$$

将式(8.3-8)和式(8.3-9)应用到本例，可以得到当α为实数时

$$K_{WY}[m,n]=\sum_{k=0}^{\infty}(1-\alpha)\alpha^k K_{WW}[m,n-k]$$

和

$$K_{YY}[m,n]=\sum_{l=0}^{\infty}(1-\alpha)\alpha^l K_{WY}[m-l,n]$$

由此结合起来得到

$$K_{YY}[m,n]=\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(1-\alpha)^2\alpha^k\alpha^l K_{WW}[m-l,n-k]$$

现在假设输入序列$W[n]$的协方差函数为

$$K_{WW}[m,n]=\sigma_W^2\delta[m-n],\quad m,n\geqslant 0$$

则输出的协方差函数可按以下步骤计算：

$$\begin{aligned}K_{YY}[m,n]&=\sum_{k=0}^{n}(1-\alpha)^2\alpha^k\alpha^{(m-n)+k}\sigma_W^2,\qquad 对\ m\geqslant n\geqslant 0\\&=\alpha^{(m-n)}(1-\alpha)^2\sum_{k=0}^{n}\alpha^{2k}\sigma_W^2\\&=\alpha^{(m-n)}\left[(1-\alpha)^2/(1-\alpha^2)\right]\sigma_W^2(1-\alpha^{2n+2}),\quad 对\ m\geqslant n\geqslant 0\\&=[(1-\alpha)/(1+\alpha)]\alpha^{|m-n|}\sigma_W^2(1-\alpha^{2\min(m,n)+2}),\ 对所有\ m,n\geqslant 0\end{aligned}$$

其中最后一步是由于(m,n)所要求的对称性。注意项$\alpha^{2\min(m,n)+2}$是瞬态的，因为$|\alpha|<1$，所以当$n\to\infty$时会逐渐消失，因此得到渐近稳态解

$$K_{YY}[m,n]=\left(\frac{1-\alpha}{1+\alpha}\right)\sigma_W^2\ \alpha^{|m-n|},\quad m,n\to\infty$$

为移不变协方差函数。如果均值函数$\mu_Y[n]$也趋向于常数，则随机序列$Y[n]$称为渐近WSS。在下一节将详细讨论WSS随机序列。

除此以外还有另一种求解方法，可以取式(8.3-12)的期望直接得到输出均值序列的递归

① 对上述方程的两端取Z变换，注意到冲激序列的Z变换为1，得到$H(z)=(1-\alpha)/(1-\alpha z^{-1})$。然后利用$Z$逆变换，即可得到上述$h[n]$(要了解$Z$逆变换，请参阅附录A)。

方程，然后可以采用8.2节的方法求解

$$\mu_Y[n]=\alpha\mu_Y[n-1]+(1-\alpha)\mu_W[n],\qquad n\geqslant 0$$

且有渐近初始条件。例如，如果$\mu_Y[-1]=0$且$\mu_W[n]=\mu_W$为给定常数，则解为

$$\mu_Y[n]=(1-\alpha^{n+1})\mu_W u[n]$$

也可以采用此方法计算输入和输出之间的互相关函数。首先，结合式(8.3-12)，然后乘以$W[m]$，最后求期望，当α为实数时

$$R_{WY}[m,n]=\alpha R_{WY}[m,n-1]+(1-\alpha)R_{WW}[m,n] \tag{8.3-13}$$

也可以用R_{WY}直接求R_{WW}。可以将式(8.3-12)表示为m的函数得到输出相关函数R_{YY}的偏差分方程，乘以$Y^*[n]$，再求其期望值，则得到

$$R_{YY}[m,n]=\alpha R_{YY}[m-1,n]+(1-\alpha)R_{WY}[m,n] \tag{8.3-14}$$

这两个差分方程可以用8.2节的方法求解，因为都可以看成一维差分方程，其中一个参数为常系数，另一个参数充当附加参数。因此可以作为n的函数对每个连续的m求解式(8.3-13)。

8.4 WSS随机序列

本节将假设要讨论的随机序列都是WSS，即

$$(1)\qquad E\{X[n]\}=\mu_X,\ \text{为常量}$$

$$(2)\quad R_{XX}[k+m,k]=E\{X[k+m]X^*[k]\}=R_{XX}[m]$$

对于二阶情形，有$E\{|X[n]|^2\}<\infty$。

平稳随机序列的自相关函数的一些重要性质列举如下。它们对于协方差函数也成立，因为它们恰是中心化随机序列$X_c[n]\triangleq X[n]-\mu_X$的自相关函数。

1. 对任意m，$|R_{XX}[m]|\leqslant R_{XX}[0]\geqslant 0$，对于$X[n]$为实数的情况，这可以直接从$E\{|X[m]-X[0]|^2\}\geqslant 0$得到，若$X[n]$非实数，则可以使用施瓦茨不等式[参见式(4.3-15)]得到。
2. $|R_{XY}[m]|\leqslant\sqrt{R_{XX}[0]R_{YY}[0]}$，由施瓦茨不等式导出。
3. $R_{XX}[m]=R_{XX}^*[-m]$，因为$R_{XX}[m]=E\{X[m+l]X^*[l]\}=E\{X[l]X^*[l-m]\}=E^*\{X[l-m]X^*[l]\}=R_{XX}^*[-m]$。
4. 对于所有$N>0$及所有复数a_1，a_2，…，a_N，必有

$$\sum_{n=1}^{N}\sum_{k=1}^{N}a_n a_k^* R_{XX}[n-k]\geqslant 0$$

性质4为自相关函数的半正定性。它是函数作为随机序列的有效自相关函数的充分必要条件。通常很难直接利用性质4去验证一个函数是否为有效的自相关函数。但是，很快将介绍一种等价的频域函数，称为功率谱密度，可实现简单的有效性测试。

对于WSS随机序列和通过卷积描述的LSI系统的情况，上一节得到的许多输入输出关系将表现出异常简单的形式。例如，由

$$Y[n] = \sum_{k=-\infty}^{+\infty} h[n-k]X[k]$$

可以得到

$$\begin{aligned}R_{XY}[m,n] &= E\{X[m]Y^*[n]\}\\ &= \sum_{k=-\infty}^{+\infty} h^*[n-k]E\{X[m]X^*[k]\}\\ &= \sum_{k=-\infty}^{+\infty} h^*[n-k]R_{XX}[m-k]\\ &= \sum_{k=-\infty}^{+\infty} h^*[-l]R_{XX}[(m-n)-l], \quad \text{并且 } l \triangleq k-n\end{aligned}$$

若输入随机序列 $X[n]$ 是 WSS。因此输入输出互相关函数 $R_{XY}[m,n]$ 是移不变的，可以利用单变量互相关函数 $R_{XY}[m] \triangleq R_{XY}[m,0]$ 写为单变量的相关函数 $R_{XX}[m]$ 的形式

$$\begin{aligned}R_{XY}[m] &= \sum_{l=-\infty}^{+\infty} h^*[-l]R_{XX}[m-l]\\ &= h^*[-m] * R_{XX}[m]\end{aligned}$$

类似地，回顾

$$\begin{aligned}R_{YY}[n+m,n] &\triangleq E\{Y[n+m]Y^*[n]\}\\ &= \sum_{k=-\infty}^{+\infty} h[k]E\{X[n+m-k]Y^*[n]\}\\ &= \sum_{k=-\infty}^{+\infty} h[k]R_{XY}[m-k]\\ &= h[m] * R_{XY}[m]\end{aligned}$$

可见输出的自相关函数也是移不变的，因此利用单参数自相关函数 $R_{YY}[m] \triangleq R_{YY}[m,0]$ 可以得到

$$R_{YY}[m] = h[m] * R_{XY}[m]$$

综合两个方程，得到

$$\begin{aligned}R_{YY}[m] &= h[m] * h^*[-m] * R_{XX}[m]\\ &= (h[m] * h^*[-m]) * R_{XX}[m] \qquad (8.4\text{-}1)\\ &= g[m] * R_{XX}[m], \quad \text{并且 } g[m] \triangleq h[m] * h^*[-m]\end{aligned}$$

其中 $g[m]$ 有时称为自相关冲激响应(AIR)。注意，如果输入随机序列是 WSS 且独立，则其自相关函数将是一个正常数乘以 $\delta[m]$，若以此常数为单位，得到的输出自相关函数将等于 $g[m]$ 本身。因此，$g[m]$ 必须满足自相关函数的所有性质，即 $g[l] = g^*[-l]$，对所有 l 都有 $g[0] \geqslant g[l]$，以及半正定性。自相关冲激响应 g 仅依赖于 LSI 系统的冲激响应 h，在没有其他信息时，难以由 g 唯一地确定 h。在天文学、结晶学以及其他一些领域，由 AIR 估计 h 的问题是一个非常重要的问题，被赋予了许多名称，包括相位恢复和去卷积。

例 8.4-1 (冲激响应)一般不能由 AIR 计算冲激响应。为了说明这一点，首先取 $g[m]$ 的傅里叶变换得到 $G(\omega)=H(\omega)H^*(\omega)=|H(\omega)|^2$。注意到 $|H(\omega)|=\sqrt{G(\omega)}$。因此 $H(\omega)$ 的相位在 AIR 中丢失，但是其幅度保留了下来。通常存在某些可用的信息来限定或确定相位，例如图像应用中 $h[n]$ 的支集，或者时基信号的因果性。感兴趣的读者有许多学术论文可以参考，如参考文献[8-6]。

例 8.4-2 (利用冲激响应分析边沿检测器的相关函数)例 8.3-2 的边沿检测器的线性变换为

$$Y[n]=L\{X[n]\}\triangleq X[n]-X[n-1]$$

冲激响应为 $h[n]=\delta[n]-\delta[n-1]$ 的 LSI 操作，其自相关函数为 $R_{XX}[m]=a^{|m|}$，其中 $|a|<1$。则很容易计算其 AIR 为

$$\begin{aligned}g[m]&=h[m]*h[-m]\\&=(\delta[m]-\delta[m-1])*(\delta[-m]-\delta[-m-1])\\&=(\delta[m]-\delta[m-1])*(\delta[m]-\delta[m+1])\\&=\delta[m]-\delta[m-1]-\delta[m+1]+\delta[m]\\&=2\delta[m]-\delta[m-1]-\delta[m+1]\end{aligned}$$

由此，对于此 WSS 可以计算输出的自相关函数为

$$\begin{aligned}R_{YY}[m]&=g[m]*R_{XX}[m]\\&=(2\delta[m]-\delta[m-1]-\delta[m+1])*a^{|m|}\\&=2a^{|m|}-a^{|m-1|}-a^{|m+1|},\qquad -\infty<m<+\infty\end{aligned}$$

这和例 8.3-2 的结果是一致的，在例 8.3-2 中画出了 $a=0.7$ 的图形。

功率谱密度

功率谱密度(psd)定义为 WSS 随机序列 $X[n]$ 的单参数离散时间自相关函数的 FT(参见定义 8.2-3)

$$S_{XX}(\omega)\triangleq\sum_{m=-\infty}^{+\infty}R_{XX}[m]\exp(-\mathrm{j}\omega m),\qquad -\pi\leqslant\omega\leqslant+\pi \tag{8.4-2}$$

现取式(8.4-1)的 FT，得到以下由 WSS 随机序列激励 LSI 系统重要的 psd 输入输出关系：

$$S_{YY}(\omega)=|H(\omega)|^2S_{XX}(\omega)=G(\omega)S_{XX}(\omega) \tag{8.4-3}$$

其中的几个频域量是离散时间傅里叶变换。式(8.4-3)是 WSS 随机序列理论的核心结论，原因在于它使得由输入 psd 和传输函数的幅度知识直接计算输出的 psd 成为可能。利用 IFT，可以计算自相关函数如下：

$$R_{XX}[m]=\mathrm{IFT}\,\{S_{XX}(\omega)\}=\frac{1}{2\pi}\int_{-\pi}^{+\pi}S_{XX}(\omega)\mathrm{e}^{\mathrm{j}\omega m}\mathrm{d}\omega$$

由此，知道了 psd 也就知道了自相关函数。

至于功率谱密度名称的由来，注意 $R_{XX}[0]=E\{|X[n]|^2\}$ 是 $X[n]$ 的统计平均功率，因此根据上面的关系，可以看出

$$E\{|X[n]|^2\}=R_{XX}[0]=\frac{1}{2\pi}\int_{-\pi}^{+\pi}S_{XX}(\omega)\mathrm{d}\omega$$

因此 psd 在频率区间$[-\pi, +\pi]$上的积分平均实际上就是平均功率。进一步地，考虑 WSS 随机序列 $X[n]$输入一个包含窄带滤波器 $H(\omega)$的 LSI 系统，其带宽 2ϵ 非常小，中心频率为 ω_o，$|\omega_o|<\pi$，通带增益为 1。记 $S_{XX}(\omega)$为输入 psd，对于输出统计平均功率，近似地有

$$R_{YY}[0]=\frac{1}{2\pi}\int_{\omega_o-\epsilon}^{\omega_o+\epsilon}S_{XX}(\omega)\,\mathrm{d}\omega\approx S_{XX}(\omega_0)\frac{\epsilon}{\pi}$$

由此可见 $S_{XX}(\omega)$可以看成平均功率在频域的密度函数。

一些 psd 的重要性质列举如下：

1. 函数 $S_{XX}(\omega)$为实数值，因为 $R_{XX}[m]$是共轭对称的。
2. 若 $X[n]$是实值随机序列，则 $S_{XX}(\omega)$是 ω 的偶函数。
3. 无论 $X[n]$取实数值还是复数值，对任何 ω，函数 $S_{XX}(\omega)\geqslant 0$。
4. 若对某个有限整数 $N>0$，当$|n|\geqslant N$ 时都有 $R_{XX}[m]=0$(即其具有有限支集)，则 $S_{XX}(\omega)$是 ω 的解析函数。这意味着给定单个点 ω_o 处的取值及其所有导数，$S_{XX}(\omega)$可以表示为一个泰勒级数。

因为 $S_{XX}(\omega)$是一个(自相关)序列的傅里叶变换，所以它是周期性的，周期为2π。这是为什么恢复自相关函数的傅里叶反变换，只在$[-\pi, +\pi]$上求积分，即其主区间。还可以定义两个联合 WSS 随机序列的互相关函数的傅里叶变换

$$S_{XY}(\omega)\triangleq\sum_{m=-\infty}^{+\infty}R_{XY}[m]\exp(-\mathrm{j}\omega m),\qquad -\pi\leqslant\omega\leqslant+\pi$$

称为随机序列 X 和 Y 之间的互功率谱密度。通常，此互功率谱密度可以是复数、负数以及非对称。其主要用途是作为计算 psd 的中间步骤。

psd 的物理意义

根据其名称可以判断 psd 应该和随机信号傅里叶变换的振幅平方的某种平均值有关。因为 WSS 随机信号 $X[n]$在所有时间具有常数平均功率 $R_{XX}[0]$，所以不能定义其 FT；但是可以定义其变换量

$$X_N(\omega)\triangleq \mathrm{FT}\{w_N[n]X[n]\}$$

其中用到矩形窗口函数

$$w_N[n]\triangleq\begin{cases}1, & |n|\leqslant N\\ 0, & \text{其他}\end{cases}$$

由此求幅度平方$|X_N(\omega)|^2$ 的期望值，并且除以 $2N+1$ 得到

$$\begin{aligned}\frac{1}{2N+1}E\{|X_N(\omega)|^2\}&=\frac{1}{2N+1}E\left\{\sum_{k=-N}^{+N}\sum_{l=-N}^{+N}X[k]X^*[l]\exp(-\mathrm{j}\omega k)\exp(+\mathrm{j}\omega l)\right\}\\&=\frac{1}{2N+1}\sum_{k=-N}^{+N}\sum_{l=-N}^{+N}E\{X[k]X^*[l]\}\exp(-\mathrm{j}\omega k)\exp(+\mathrm{j}\omega l)\\&=\frac{1}{2N+1}\sum_{k=-N}^{+N}\sum_{l=-N}^{+N}R_{XX}[k-l]\exp[-\mathrm{j}\omega(k-l)]\\&=\sum_{m=-2N}^{+2N}R_{XX}[m]\left(1-\frac{|m|}{2N+1}\right)\exp(-\mathrm{j}\omega m)\end{aligned}$$

其中最后一行是由于在$(k,\ l)$平面中沿着$(2N+1)\times(2N+1)$点构成的矩形的对角线$k-l=m$上，$R_{XX}[k-l]$为常数。

因为随着$|m|\to\infty$时$|R_{XX}[m]|\to 0$，对于傅里叶变换，这是保证$S_{XX}(\omega)$存在的必然要求，所以当$N\to\infty$，三角函数$\left(1-\frac{|m|}{2N+1}\right)$的影响越来越微弱。实际上，如果假设$|R_{XX}[m]|$衰减足够快，足以保证$\sum_{m=-\infty}^{+\infty}|m||R_{XX}[m]|<\infty$，则有

$$S_{XX}(\omega)=\lim_{N\to\infty}\frac{1}{2N+1}E\{|X_N(\omega)|^2\} \tag{8.4-4}$$

通俗地讲就是，加窗随机序列在频率ω上的统计平均功率由功率谱密度函数$S_{XX}(\omega)$确定。注意前面没有提到过随机变量$\frac{1}{2N+1}|X_N(\omega)|^2$的方差，只说过其平均值收敛到psd。在学习频谱估计时(参见11.6节)，会证明随着N变大，方差不会减小，因此$\frac{1}{2N+1}|X_N(\omega)|^2$在取平均之前，不能作为好的psd的估计。在统计学(参见第6章)中，我们说$\frac{1}{2N+1}|X_N(\omega)|^2$不是$S_{XX}(\omega)$的一致估计。

例8.4-3 下面是计算例8.1-16中$p=0.8$, 0.5, 0.2的有记忆随机序列的psd的MATLAB源程序。

```
function[psd1,psd2,psd3]=psdmarkov2(N,p1,p2,p3)}
mc1=0*ones(1,N);
mc2=0*ones(1,N);
mc3=0*ones(1,N);
for i=1:N
 mc1(i)=0.25*(((-1)*(2*p1-1))^(i-1));% The (-1)^(i-1) factor shifts the
spectrum to yield
 mc2(i)=0.25*(((-1)*(2*p2-1))^(i-1));%an even function of frequency.
Otherwise
 mc3(i)=0.25*(((-1)*(2*p3-1))^(i-1));%the highest frequency
components appear

end
x=linspace(-pi,pi,N);%at pi and the lowest at 2*pi.
psd1=abs(fft(mc1));
psd2=abs(fft(mc2));
psd3=abs(fft(mc3));
plot(x,psd1,x,psd2,x,psd3)
title('Power spectral density (psd) of random sequences with memory')
xlabel('radian frequency')
ylabel('psd value')
end
```

相应的曲线参见图8.4-1。

例8.4-4 平稳随机序列$X[n]$具有功率谱密度$S_{XX}(\omega)=N_0w(3\omega/4\pi)$，其中矩形窗函数$w$为

$$w(x)\triangleq\begin{cases}1, & |x|\leqslant 1/2\\ 0, & \text{其他}\end{cases}$$

要求输出随机序列$Y[n]$的psd为$S_{YY}(\omega)=N_0w(\omega/\pi)$。现给定一LSI系统(不一定是因果的)，其冲激响应为$h[n]$。应该使用下面哪个冲激响应[注意$\text{sinc}(x)\triangleq\sin(\pi x)/\pi x$]?

(a) $2\text{sinc}(n/2)$

(b) $\frac{1}{2}\text{sinc}((n-10)/2)$

(c) $1.5\mathrm{e}^{-|n|}u[n]$

(d) $u[n+2]-u[n-2]$

(e) $(1-|n|)w(n/2)$

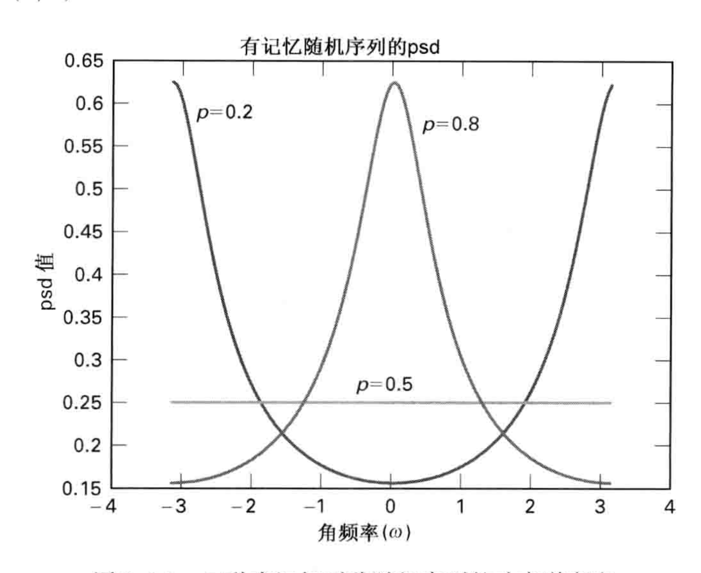


图 8.4-1　三种有记忆平稳随机序列的功率谱密度

解　显然需要的是振幅为 $|H(\omega)|=w(\omega/\pi)$ 的转移函数 $H(\omega)$。这样立刻可以排除选项(c)到(e)，因为它们的傅里叶变换在任何频带都不是常数振幅。因为 $w(\omega/\pi)$ 的 IFT 为 $\frac{1}{2}\text{sinc}(n/2)$，所以选择(b)，因为 10 个样本的延迟并不影响其幅度为 $|H(\omega)|$。

表 8.4-1 对随机序列输入输出关系进行了有益的总结。

表 8.4-1　WSS 序列和线性系统输入输出关系

随机序列	输出均值
$Y[n]=h[n]*X[n]$	$\mu_Y=H(0)\mu_X$
互相关函数	互功率谱密度
$R_{XY}[m]=R_{XX}[m]*h^*[-m]$	$S_{XY}(\omega)=S_{XX}(\omega)H^*(\omega)$
$R_{YX}[m]=h[m]*R_{XX}[m]$	$S_{YX}(\omega)=H(\omega)S_{XX}(\omega)$
$R_{YY}[m]=R_{YX}[m]*h^*[-m]$	$S_{YY}(\omega)=S_{YX}(\omega)H^*(\omega)$
自相关函数	功率谱密度
$R_{YY}[m]=h[m]*h^*[-m]*R_{XX}[m]$	$S_{YY}(\omega)=\|H(\omega)\|^2S_{XX}(\omega)$
$=g[m]*R_{XX}[m]$	$=G(\omega)S_{XX}(\omega)$
输出功率及方差	
$E\{\|Y[n]\|^2\}=R_{YY}[0]=\frac{1}{2\pi}\int_{-\pi}^{+\pi}\|H(\omega)\|^2S_{XX}(\omega)\mathrm{d}(\omega)$	
$\sigma_Y^2=R_{YY}[0]-\|\mu_Y\|^2$	

随机序列合成与离散时间模拟

接下来考虑如何求出适当的转移函数 $H(\omega)$ 来产生具有指定 psd 或相关函数的随机序列。回顾式(8.2-2)，为方便起见，这里重复一下：

$$y[n]=\sum_{k=1}^{M}a_k y[n-k]+\sum_{k=0}^{N}b_k x[n-k] \tag{8.4-5}$$

其中的系数都是实数。其转移函数 $H(\omega)$ 由下式给出：

$$H(\omega)=\frac{Y(\omega)}{X(\omega)}=\frac{B(\omega)}{A(\omega)}$$

其中 $B(\omega)\triangleq\sum_{k=0}^{N}b_k\mathrm{e}^{-\mathrm{j}\omega k}$ 及 $A(\omega)\triangleq 1-\sum_{k=1}^{M}a_k\mathrm{e}^{-\mathrm{j}\omega k}$。当输入功率为 $E\{|W[n]|^2\}=\sigma_W^2$，白噪声序列 $W[n]$，输出的 psd $S_{YY}(\omega)$ 如下：

$$S_{YY}(\omega)=\frac{|B(\omega)|^2}{|A(\omega)|^2}\sigma_W^2=\frac{B(\omega)B^*(\omega)}{A(\omega)A^*(\omega)}\sigma_W^2 \tag{8.4-6}$$

又根据 $z=\mathrm{e}^{\mathrm{j}\omega}$ 时 $B(z)\triangleq B(\omega)$，以及同样的定义 $A(z)$，$H(z)$，另外根据 $\mathrm{e}^{-\mathrm{j}\omega}=z^{-1}$ 和 $B^*(\mathrm{e}^{\mathrm{j}\omega})=B(\mathrm{e}^{-\mathrm{j}\omega})$，可以得到具有实系数①的 LCCDE

$$S_{YY}(z)=\frac{B(z)B(z^{-1})}{A(z)A(z^{-1})}\sigma_W^2=H(z)H(z^{-1})\sigma_W^2 \tag{8.4-7}$$

到目前为止，z 限制在单位圆上。为了进一步分析，有必要将式(8.4-7)扩展到整个 z 平面。

最后的一步称为解析延拓，只是用来求出在单位圆 $z=\mathrm{e}^{\mathrm{j}\omega}$ 上的取值与给定 psd 一致的 z 的有理函数。

给定任意有理函数 $S_{XX}(z)$，即在有限 z 平面上存在有限个极点和零点的函数，通过定义 $H(z)$ 的单位圆中 $\{|z|<1\}$ 包含所有的极点和零点，可以求出类似式(8.4-7)的谱分解，而且 $H(z^{-1})$ 中必然有所有的极点和零点都在单位圆外面 $\{|z|>1\}$。

例 8.4-5 考虑 psd

$$S_{XX}(\omega)=\frac{\sigma_W^2}{1-2\rho\cos\omega+\rho^2}\qquad 且\ |\rho|<1$$

首先要求扩展 $S_{XX}(\omega)$ 到整个 z 平面。由 $\cos\omega=\frac{1}{2}(\mathrm{e}^{+\mathrm{j}\omega}+\mathrm{e}^{-\mathrm{j}\omega})$，其可以扩展为 $\frac{1}{2}(z+z^{-1})$ 且满足实值随机序列的对称条件 $S_{XX}(z)=S_{XX}(z^{-1})$。因此

$$\begin{aligned}S_{XX}(z)&=\frac{\sigma_W^2}{1-\rho(z+z^{-1})+\rho^2},\qquad |\rho|<|z|<\frac{1}{|\rho|}\\&=\frac{\sigma_W^2}{(1-\rho z)(1-\rho z^{-1})}\\&=\sigma_W^2H(z)H(z^{-1})\end{aligned}$$

对于收敛域 $|\rho|<|z|$，$H(z)=\dfrac{1}{1-\rho z^{-1}}$。

---

① 只有在类似这里的情况，冲激响应的系数取实数。这是因为这里分子及分母的系数都是实数。

因为 $|\rho|<1$，收敛域(ROC)包含单位圆，因此 H 是稳定因果的。实际上，具有 $h[n]=\rho^n u[n]$ 的系统将由独立的序列产生 $S_{XX}(\omega)$。

如果单位圆上出现零点，则其必然是偶数阶的，否则可以很容易证明$S_{XX}(\mathrm{e}^{\mathrm{j}\omega})$必然经过零点，因此在其附近为负数。所以，可以将一半的零点分配给 $H[n]$，另一半分配给 $H(z^{-1})$。因为在 $H[z]$ 的单位圆中只有极点，所以它是 BIBO 稳定的[8-5]。除非在单位圆上出现一个零点，否则其逆也是稳定的。另一方面所有的极点都在 $H(z^{-1})$ 的单位圆外面，因此从非因果的角度它是稳定的。记单位圆中极点幅度最大值为 $p_{\max}$，从而 $S_{XX}(z)$ 是解析的，即在环形收敛域 $\{p_{\max}<|z|<1/p_{\max}\}$ 中不存在奇异点。

继续上述过程，可以得到被白噪声 $W[n]$ 激励，系统函数 $H(z)$ 将生成 psd 为 $S_{XX}(\omega)$ 的随机序列 $X[n]$。这可以作为计算机上离散时间模拟的基础。白随机序列 $W[n]$ 很容易利用计算机的随机数发生器得到。然后指定适当的初始条件，接着利用系统函数 $H[n]$ 的 LCCDE 递归计算 $X[n]$。

为使 X 达到高斯分布，可以将随机数发生器的输出进行变换从而使 W 达到高斯分布，该分布特性将传递到 X。经常采用的一种近似方法是根据中心极限定理，平均 6 ~ 10 次随机数调用的值，从而使 W 接近高斯分布。如果对一个非高斯随机变量进行仿真，则 X 和 W 的分布是不同的。因此后续的方法难以奏效。一种可能性是利用 LCCDE 由某些实际的数据产生$W[n]$ 的样本，然后在仿真中使用此分布的 $W[n]$。

例 8.4-6 （匹配给定相关值）为了仿真平均功率为 $R_{XX}[0]=\sigma^2$，最近邻相关为 $R_{XX}[1]=\rho\sigma^2$ 的零均值随机序列，需要找到一阶随机差分方程的系数以满足这些值。因此考虑

$$X[n]=aX[n-1]+bW[n] \tag{8.4-8}$$

其中 $W[n]$ 是单位功率零均值白噪声源。计算其冲激响应，得到

$$h[n]=ba^n u[n]$$

以及相应的系统函数

$$H(z)=\frac{b}{1-az^{-1}}$$

由于均值为零，计算式(8.4-8)中输出 $X[n]$ 的协方差

$$\begin{aligned}
K_{XX}[m] &= h[m]*h[-m]*K_{WW}[m] \\
&= h[m]*h[-m] \\
&= b^2\left(a^m u[m]\right)*\left(a^{-m}u[-m]\right) \\
&= b^2\sum_{k=-\infty}^{+\infty} a^k u[k] a^{m+k} u[m+k] \\
&= b^2 a^m \sum_{k=\max(0,-m)}^{+\infty} a^{2k} \\
&= \frac{b^2}{1-a^2}a^{|m|}, \qquad -\infty<m<+\infty
\end{aligned}$$

从 $m=0$ 和 $m=1$ 处的要求，需要满足

$$K_{XX}[0]=\sigma^2=b^2/(1-a^2)$$

$$K_{XX}[1]=\rho\sigma^2=ab^2/(1-a^2)$$

由此

$$a=\rho \text{ 和 } b^2=\sigma^2(1-\rho^2)$$

为计算最终的 psd，需利用式(8.4-4)，得到

$$\begin{aligned}S_{XX}(\omega)&=\frac{b^2}{|1-a\mathrm{e}^{-\mathrm{j}\omega}|^2}\\&=\frac{\sigma^2(1-\rho^2)}{1-2\rho\cos\omega+\rho^2}\end{aligned}$$

例 8.4-7 (抽取与内插)令 $X[n]$ 为 WSS 随机序列。考虑在许多信号处理应用中采用的抽取和内插操作对其平稳性和 psd 的影响。

抽取

令 $Y[n]\triangleq X[2n]$，称为因子 2 的抽取，即丢弃每个索引为奇数的样本 $X[n]$(如图 8.4-2 所示)。很容易计算均值函数 $\mu_Y[n]\triangleq E\{Y[n]\}=E\{X[2n]\}=\mu_X[2n]=\mu_X$ 为常数。对于相关函数

$$\begin{aligned}R_{YY}[n+m,n]&=E\{X[2n+2m]X^*[2n]\}\\&=R_{XX}[2n+2m,2n]\\&=R_{XX}[2m]\end{aligned}$$

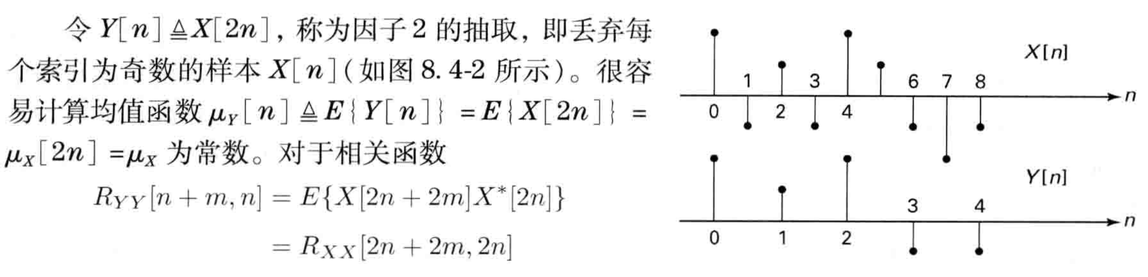


图 8.4-2 在抽取中 $X[n]$ 每两个值丢弃一个

可见原随机序列 $X[n]$ 的 WSS 性质在抽取随机序列中得以保留。$Y[n]$ 的 psd 可以如下：

$$\begin{aligned}S_{YY}(\omega)&\triangleq\sum_{m=-\infty}^{+\infty}R_{YY}[m]\exp[-\mathrm{j}\omega m]\\&=\sum_{m=-\infty}^{+\infty}R_{XX}[2m]\exp[-\mathrm{j}\omega m]\\&=\sum_{m\text{ 为偶数}}R_{XX}[m]\exp\left[-\mathrm{j}\frac{\omega}{2}m\right]=\sum_{m\text{ 为偶数}}R_{XX}[m]\exp\left[-\mathrm{j}\frac{\omega}{2}m\right](-1)^m\end{aligned}$$

现定义 $A_e\triangleq S_{YY}(\omega)$，$A_o\triangleq\sum_{m\text{ 为奇数}}R_{XX}[m]\exp\left[-\mathrm{j}\frac{\omega}{2}m\right]$。则显然 $A_e+A_o=S_{XX}\left(\frac{\omega}{2}\right)$ 及 $A_e-A_o=S_{XX}\left(\frac{\omega-2\pi}{2}\right)$，由此可得

$$S_{YY}(\omega)=\frac{1}{2}\left[S_{XX}\left(\frac{\omega}{2}\right)+S_{XX}\left(\frac{\omega-2\pi}{2}\right)\right]$$

其表现为高频分量的衰减[8-5]。

内插

对于因子 2 的内插，需要做与抽取相反的事。首先通过下式进行扩展：

$$X_e[n]\triangleq\begin{cases}X[\frac{n}{2}], & n=\text{偶数}\\0, & n=\text{奇数}\end{cases}$$

因为零值的插入，得到的扩展随机序列显然是非平稳的，如图 8.4-3 所示。形式上，$X_e[n]$ 的 psd 不存在，因为 psd 只对 WSS 序列(参见图 8.4-4)有定义。对于一大类称为周期平稳(参见 9.6 节)的随机序列和随机过程①，$X_e[n]$ 即属于此列，都会碰到这种问题。这种序列可以方便地转化为 WSS，只要对起始时间随机化然后对起始时间取平均[参见式(9.6-1)]。但是这里要利用式(8.4-4)计算功率谱密度，且对于周期平稳波形也可以采用。由此，记

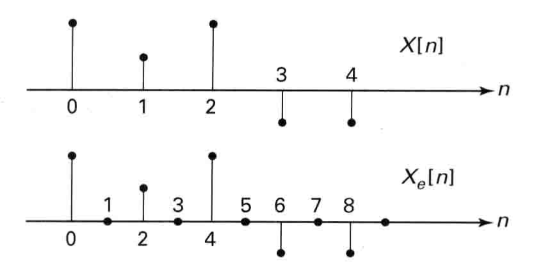


图 8.4-3　在内插中，扩展步骤在 $X[n]$ 序列的相邻值之间插入零值，得到扩展序列 $X_e[n]$

$$E\{|Y_N(\omega)|^2\} = E\left\{\left|\sum_{n=-N}^{N} X_e[n]\mathrm{e}^{-\mathrm{j}\omega n}\right|^2\right\}$$

且取 $N\to\infty$ 时 $E\{|X_e^{(N)}(\omega)|^2\}/(2N+1)$ 的极限。这可以看成随机序列 $X_e[n]$ 的 psd，$S_{X_eX_e}(\omega)$。若进一步化简并假设 $R_{XX}[m]$ 是绝对可加的，则 $S_{X_eX_e}(\omega)=\frac{1}{2}S_{XX}(2\omega)$。对于扩展步骤的深入讨论，可参考习题 8.58 和习题 8.59。

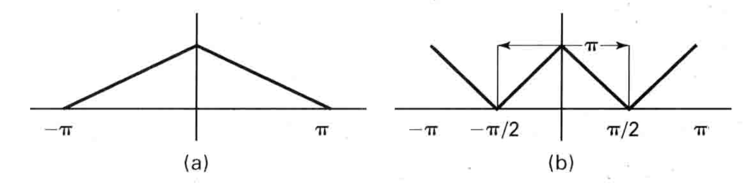


图 8.4-4　(a) 原始 $X[n]$ 的 psd；(b) $X_e[n]$ 的 psd(未按比例)。注意功率谱密度从第二周期泄漏到主周期。定义在 $[-\pi/2,\pi/2]$ 上的理想低通滤波器可以消除第二周期信号的影响

接下来将 $X_e[n]$，有时称为 $X[n]$ 的过采样版本输入到带宽为 $\left[-\frac{\pi}{2},\ +\frac{\pi}{2}\right]$，通带增益为 2 的低通滤波器，产生“理想的”内插输出 $Y[n]$ 为

$$Y[n] \triangleq h[n] * X_e[n]$$

此滤波器的冲激响应为

$$h[n] = \frac{\sin(\pi n/2)}{(\pi n/2)}$$

由此

$$\begin{aligned} Y[n] &= \sum_{k=-\infty}^{+\infty} X_e[k]\frac{\sin(n-k)\pi/2}{(n-k)\pi/2} \\ &= \sum_{k=-\infty}^{+\infty} X[k]\frac{\sin(n-2k)\pi/2}{(n-2k)\pi/2} \end{aligned}$$

首先计算 $Y[n]$ 的均值函数

① 随机过程是第 9 章将要讨论的连续时间随机波形。

$$\begin{aligned}\mu_Y[n] &\triangleq E\{Y[n]\}\\ &= E\left\{\sum_{k=-\infty}^{+\infty} X[k]\frac{\sin(n-2k)\pi/2}{(n-2k)\pi/2}\right\}\\ &= \sum_{k=-\infty}^{+\infty} \mu_X[k]\frac{\sin(n-2k)\pi/2}{(n-2k)\pi/2}\\ &= \mu_X\sum_{k=-\infty}^{+\infty} \frac{\sin(n-2k)\pi/2}{(n-2k)\pi/2}\end{aligned}$$

最后一步是因为μ_X是常数。利用采样定理可以证明无限和为1，因此$\mu_Y[n]=\mu_X$。为此，记任意带限函数$g(t)$，采样周期$T=2$的采样定理表示为[8-5]

$$g(t)=\sum_{k=-\infty}^{+\infty} g(2k)\frac{\sin(t-2k)\pi/2}{(t-2k)\pi/2} \tag{8.4-9}$$

接着只要在带宽为零的带限函数$g(t)=1$中选择$t=n$，即可得

$$1=\sum_{k=-\infty}^{+\infty}\frac{\sin(n-2k)\pi/2}{(n-2k)\pi/2}$$

为方便后面引用，定义$h(t)\triangleq\dfrac{\sin\pi t/2}{\pi t/2}$。为了求相关函数，进一步计算

$$\begin{aligned}R_{YY}[n+m,n] &= E\{Y[n+m]Y^*[n]\}\\ &= \sum_{k_1,k_2=-\infty}^{+\infty} E\{X[k_1]X^*[k_2]\}h[n+m-2k_1]h[n-2k_2]\\ &= \sum_{l_1=\text{偶数}} R_{XX}[l_1]\sum_{l_2=\text{偶数}} h[n+m-l_1-l_2]h[n+l_1-l_2]\\ &\quad+\sum_{l_1=\text{奇数}} R_{XX}[l_1]\sum_{l_2=\text{奇数}} h[n+m-l_1-l_2]h[n+l_1-l_2]\end{aligned}$$

其中$l_1\triangleq k_1-k_2$，$l_2\triangleq k_1+k_2$以及$l_2\pm l_1$为偶数。可以令式(8.4-9)中$g(t)=h(t)$且令$t=m$计算其和

$$\sum_{l_2=\text{偶数或奇数}} h[n+m-l_1-l_2]h[n+l_1-l_2]$$

可以看到求和结果，无论对偶数还是奇数，都等于$h[m-2l_1]$。因此

$$R_{YY}[m+n,n]=R_{YY}[m]=\sum_{l_1}R_{XX}[l_1]h[m-2l_1]$$

由此可见$Y[n]$是WSS，且$R_{YY}[m]$是$R_{XX}[m]$的内插，即

$$\begin{aligned}R_{YY}[2m] &= \sum_{l_1}R_{XX}[l_1]h[2m-2l_1]\\ &= R_{XX}[2m]\end{aligned}$$

且由此计算psd如下：

$$
\begin{aligned}
S_{YY}(\omega) &= \sum_m R_{YY}[m]\mathrm{e}^{-\mathrm{j}\omega m} \\
&= \sum_m \sum_{l_1} R_{XX}[l_1]h[m-2l_1]\mathrm{e}^{-\mathrm{j}\omega m} \\
&= \sum_{l_1} R_{XX}[l_1]\sum_m h[m-2l_1]\mathrm{e}^{-\mathrm{j}\omega m} \\
&= \sum_{l_1} R_{XX}[l_1]H(\omega)\mathrm{e}^{-\mathrm{j}2\omega l_1} \\
&= H(\omega)S_{XX}(2\omega) \\
&= \begin{cases} 2S_{XX}(2\omega), & |\omega| \leqslant \pi/2, \\ 0, & \dfrac{\pi}{2} < |\omega| \leqslant \pi \end{cases}
\end{aligned}
$$

8.5　马尔可夫随机序列

我们已经见过一些马尔可夫随机序列的例子。这种序列有时通俗地称为有记忆的或者有状态的。接下来进一步明确这些概念。首先给出定义。

定义 8.5-1　（马尔可夫随机序列）

(a) 连续马尔可夫随机序列 $X[n]$，对 $n \geqslant 0$，满足以下条件概率密度表达式：

$$f_X(x_{n+k}|x_n, x_{n-1}, \cdots, x_0) = f_X(x_{n+k}|x_n)$$

对所有 x_0，…，x_n，x_{n+k}，$n>0$，以及所有整数 $k \geqslant 1$ 都成立。

(b) 离散马尔可夫随机序列 $X[n]$，对 $n \geqslant 0$，满足以下条件概率质量函数表达式：

$$P_X(x_{n+k}|x_n, \cdots, x_0) = P_X(x_{n+k}|x_n)$$

对所有 x_0，…，x_n，x_{n+k}，$n>0$，以及所有 $k \geqslant 1$ 都成立。

对于 $k=1$，即单步的情况，上述条件都成立则是充分的，因为一般的性质可以由此构造。离散马尔可夫随机序列也称为马尔可夫链，将在下一节讨论。这里考虑连续取值的情况。

为了检验马尔可夫概念的内涵和意义，考虑随机序列 X 的 N 阶概率密度函数 $f_X(x_N, x_{N-1}, \cdots, x_0)$，反复利用条件可以得到概率的链式规则

$$f_X(x_0, x_1, \cdots, x_N) = f_X(x_0)f_X(x_1|x_0)f_X(x_2|x_1, x_0)\cdots f_X(x_N|x_{N-1}, \cdots, x_0) \tag{8.5-1}$$

将基本的单步 ($k=1$) 马尔可夫定义代入该式，得到

$$
\begin{aligned}
f_X(x_0, x_1, \cdots, x_N) &= f_X(x_0)f_X(x_1|x_0)f_X(x_2|x_1)\cdots f_X(x_N|x_{N-1}) \\
&= f_X(x_0)\prod_{k=1}^{N} f_X(x_k|x_{k-1})
\end{aligned}
$$

下面举两个满足高斯分布的连续马尔可夫随机序列的例子。

例 8.5-1　（高斯马尔可夫随机序列）令 $X[n]$ 为对 $n \geqslant 1$ 有定义的随机序列，初始概率密度为

$$f_X(x;0) \approx N(0, \sigma_0^2)$$

其中 $\sigma_0 > 0$，转移概率密度为

$$f_X(x_n|x_{n-1};n,n-1)\approx N(\rho x_{n-1},\sigma_W^2)$$

其中$|\rho|<1$且$\sigma_W>0$。需要确定$X[n]$在任意$n\geqslant 1$及其后续时间的无条件概率密度。

一般地，需要从初始密度通过积分[参见式(2.6-84)]递归地计算

$$f_X(x;n)=\int_{-\infty}^{+\infty}f_X(x|\zeta;n,n-1)f_X(\zeta;n-1)\mathrm{d}\zeta \tag{8.5-2}$$

对$n=1,2,3$。但是在本例中，已经知道无条件一阶密度是高斯的，因为式(8.5-2)中的每个概率密度函数都是高斯的，而且高斯密度在这种情况下“再现自身”；也就是，两个指数函数的乘积仍然是指数函数。因此概率密度函数$f_X(x;n)$由前两个时刻确定。首先计算其均值函数

$$\begin{aligned}\mu_X[n]&=E\{X[n]\}\\&=E[E\{X[n]|X[n-1]\}]\\&=E[\rho X[n-1]]\\&=\rho\mu_X[n-1]\end{aligned}$$

其中外面的期望是对$X[n-1]$的值求解。由此得到递归方程

$$\mu_X[n]=\rho\mu_X[n-1],\qquad n\geqslant 1$$

具有预先给定的初始条件$\mu_X[0]=0$。由此对所有n，$\mu_X[n]=0$。

同样也需要求方差函数$\sigma_X^2[n]$，在本例中恰为$E[X^2[n]]$，因为均值为0。计算得到

$$\begin{aligned}E\{X^2[n]\}&=E[E\{X^2[n]|X[n-1]\}]\\&=E[\sigma_W^2+\rho^2X^2[n-1]]\\&=\sigma_W^2+\rho^2E\{X^2[n-1]\}\end{aligned}$$

或者

$$\sigma_X^2[n]=\rho^2\sigma_X^2[n-1]+\sigma_W^2,\qquad n\geqslant 1$$

这是一个一阶差分方程，通过由初始概率密度函数给定的$\sigma_X^2[0]=\sigma_0^2$可以求解$\sigma_X^2[n]$。解为

$$\begin{aligned}\sigma_X^2[n]&=[1+\rho^2+\rho^4+\cdots+\rho^{2(n-1)}]\sigma_W^2+\rho^{2n}\sigma_0^2\\&\to\frac{1}{1-\rho^2}\sigma_W^2,\qquad n\to\infty\end{aligned}$$

例 8.5-2 (马尔可夫差分方程)考虑差分方程

$$X[n]=\rho X[n-1]+W[n]$$

其中$W[n]$是独立随机序列(参见定义8.1-2)。令$n>0$；则

$$\begin{aligned}f_X(x_n,x_{n-1},\cdots,x_0)&=f_X(x_n|x_{n-1})f_X(x_{n-1}|x_{n-2})\cdots f_X(x_1|x_0)f_W(x_0)\\&=\left(\prod_{k=1}^{n}f_W(x_k-\rho x_{k-1})\right)f_W(x_0)\end{aligned}$$

其中$x[n]=x_n$及$w[n]=w_n$分别为取自随机序列$X[n]$和$W[n]$的样本函数。显然$X[n]$是马尔可夫随机序列。如果$W[n]$是一个独立的高斯随机序列，则这就是例8.5-1的情况。否则马尔可夫序列$X[n]$将是非高斯的。

马尔可夫性质可以推广到高阶相关以及高阶差分方程，从而将直接相关概念扩展到多于一个样本距离，扩展直接依赖于多个样本实例。

定义 8.5-2　（马尔可夫-p 随机序列）令正整数 p 称为马尔可夫-p 随机序列的阶。连续马尔可夫-p 随机序列 $X[n]$，对 $n\geqslant 0$ 有定义，以下条件概率密度方程：

$$f_X(x_{n+k}|x_n,x_{n-1},\cdots,x_0)=f_X(x_{n+k}|x_n,x_{n-1},\cdots,x_{n-p+1})$$

对所有 $k\geqslant 1$ 及 $n\geqslant p$ 都成立。

回顾式(8.5-1)，可以发现随着马尔可夫阶 p 增大，用马尔可夫随机序列模拟一般随机序列的建模误差将逐步改善。

$$\begin{aligned}
&f_X(x_0,x_1,\cdots,x_N)\\
&\quad=f_X(x_0)f_X(x_1|x_0)f_X(x_2|x_1,x_0)\cdots f_X(x_N|x_{N-1},\cdots,x_0)\\
&\quad\approx f_X(x_0)f_X(x_1|x_0)f_X(x_2|x_1,x_0)\cdots f_X(x_p|x_{p-1},\cdots,x_0)\\
&\qquad\times\prod_{k=p+1}^{N}f_X(x_k|x_{k-1},\cdots,x_{k-p+1})
\end{aligned}$$

这种模拟对于一般情况都是成立的，即最近的依赖最强，比如 $X[n-1]$ 和 $X[n-2]$，而离得较远的依赖性通常可以忽略。当马尔可夫-p 模型用于信号处理中，最重要的问题是确定适当的模型阶数 p，使得原始数据的诸如联合概率密度函数式(8.5-1)之类的统计特性能够被马尔可夫-p 模型适当模拟。著名的用于噪声中失真信号的递归线性估计的卡尔曼滤波器就是基于马尔可夫模型。

ARMA 模型

有一类线性常系数差分方程模型被称为 ARMA，表示自回归移动平均。这里的输入是独立的随机序列 $W[n]$，均值 $\mu_W=0$，方差 $\sigma_W^2=1$。其 LCCDE 模型形如

$$X[n]=\sum_{k=1}^{M}a_kX[n-k]+\sum_{k=0}^{L}b_kW[n-k]$$

若模型是 BIBO 稳定且 $-\infty<n<+\infty$，则输出一 WSS 序列，其 psd 为

$$S_{XX}(\omega)=\frac{\left|\sum_{k=0}^{L}b_k\exp(-\mathrm{j}\omega k)\right|^2}{\left|1-\sum_{k=1}^{M}a_k\exp(-\mathrm{j}\omega k)\right|^2}$$

ARMA 序列不是马尔可夫的，但是当 $L=0$ 时，它是马尔可夫-M，得到的模型称为自回归的(AR)。另一方面，当 $M=0$ 时，即不存在反馈系数 a_k，则 ARMA 序列变成

$$X[n]=\sum_{k=0}^{L}d_kW[n-k]$$

则模型称为移动平均的(MA)。MA 模型经常用于估计数据窗口内的时间平均，如下例所示。

例 8.5-3　(运动时间平均)考虑 $n \geqslant 1$ 的独立随机变量 $W[n]$ 的序列。记其运动时间平均为

$$\widehat{\mu}_W[n] = \frac{1}{n}\sum_{k=1}^{n} W[k]$$

因为可以等价地将 $\hat{\mu}_W[n]$ 写为满足时变 AR 方程

$$\widehat{\mu}_W[n] = \frac{n-1}{n}\widehat{\mu}_W[n-1] + \frac{1}{n}W[n]$$

根据输入 $W[n]$ 联合独立性，得到 $\hat{\mu}_W[n]$ 是非平稳的马尔可夫随机序列①。

马尔可夫链

马尔可夫随机序列既可以取连续值，也可以取离散值，因此相应地可以被表示为概率密度函数或者概率质量函数。对于离散取值的情况，称随机序列为马尔可夫链。其应用包括：缓冲区占用、计算机网络以及连续时间马尔可夫链的离散时间仿真模型(参见第 9 章)。

定义 8.5-3　(马尔可夫链)离散时间马尔可夫链是一个随机序列 $X[n]$，其 N 阶条件概率质量函数满足

$$P_X(x[n]|x[n-1],\cdots,x[n-N]) = P_X(x[n]|x[n-1]) \tag{8.5-3}$$

对所有 n、所有 $x[k]$ 的取值以及对所有整数 $N>1$ 都成立。

$X[n]$ 在时间 n 处的值称为"状态"。这是因为当前的值，也就是时刻 n 的值确定了后续条件概率质量函数，而与 $X[n]$ 之前的取值无关。

一个特别重要的实例是，当 $X[n]$ 的取值范围有限，比如 M。$X[n]$ 的离散区间，也就是 X 取值的范围，有时被称为标签集合。标签集合通常的选择是整数 $\{1, M\}$ 或 $\{0, M-1\}$。这种马尔可夫链称为具有有限状态空间，或者简称有限状态马尔可夫链。此时，而且当随机序列是平稳的，可以用具有如下元素的矩阵 $\boldsymbol{P}$ 表示概率转移信息

$$p_{ij} = P_{X[n]|X[n-1]}(j|i) \tag{8.5-4}$$

其中 $1 \leqslant i, j \leqslant M$。矩阵 $\boldsymbol{P}$ 被称为状态转移矩阵。根据其定义有，此矩阵的所有元素非负，每一行的和为 1。通常，不失一般性，可以认为马尔可夫链起始于时刻 $n=0$。因此必须明确 $n=0$ 时刻状态初始概率的集合，即 $P_X(i;0)$，$1 \leqslant i \leqslant M$，这可以保存在初始概率向量 $\boldsymbol{p}[0]$，这是一个行向量，其元素为 $(\boldsymbol{p}[0])_i = P_X(i;0)$，$1 \leqslant i \leqslant M$。

下面的例子介绍非常有用的，在例 8.1-14 中出现过的状态转移图。

例 8.5-4　(二状态马尔可夫链)令 $M=2$；则可以总结一个二状态马尔可夫链的状态转移信息如图 8.5-1 所示。唯一需要增加的是初始概率的设置，$P_X(1;0)$ 和 $P_X(2;0)$。

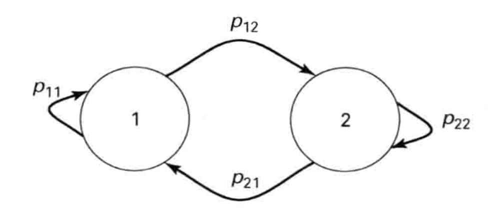


图 8.5-1　二状态马尔可夫链的状态转移图

可能的问题是：假设在时刻 4 处于状态 1，那么在时刻 6 处于状态 2 的概率是多少呢？或者给定

① 注意 $\hat{\mu}_W[n]$ 的方差随 n 递减。

时刻 3 两个状态的概率分布，那么在时刻 7 两个状态的概率分布又是多少？注意，在某个时刻在一个状态，而在若干时间之后转移到另一个状态存在几种方法或途径。这些问题的答案将包含对这些互斥输出的求和。

这里有 $\boldsymbol{M}=2$，两个元素的概率行向量 $\boldsymbol{p}[n]=(p_0[0],p_1[n])$。利用状态转移矩阵，有

$$\boldsymbol{p}[1]=\boldsymbol{p}[0]\boldsymbol{P}$$

$$\boldsymbol{p}[2]=\boldsymbol{p}[1]\boldsymbol{P}$$

$$=\boldsymbol{p}[0]\boldsymbol{P}^2$$

或者，一般地

$$\boldsymbol{p}[n]=\boldsymbol{p}[0]\boldsymbol{P}^n$$

如果存在统计稳态，将得到 $\boldsymbol{p}[\infty]=\boldsymbol{p}[\infty]\boldsymbol{P}$，其中 $\boldsymbol{p}[\infty]=\lim_{n\to\infty}\boldsymbol{p}[n]$。

记 $\boldsymbol{p}\triangleq\boldsymbol{p}[\infty]$，则有 $\boldsymbol{p}(\boldsymbol{I}-\boldsymbol{P})=\boldsymbol{0}$，包含 $M-1$ 个独立的线性方程。借助于额外的等式 $\boldsymbol{p}\boldsymbol{1}=1$，其中 $\boldsymbol{1}$ 是 M 个元素的全 1 列向量，则可以求解 $\boldsymbol{p}$ 中的 M 个值。稳定状态的存在，或者等价地渐近平稳，将依赖于状态转移矩阵 $\boldsymbol{P}$ 的特征值。

例 8.5-5　(非对称二状态马尔可夫链)下面考虑一个二状态、非对称马尔可夫链(AMC)的例子，状态标签为 0 和 1，状态转移矩阵为

$$\boldsymbol{P}=\begin{bmatrix}p_{00} & p_{01}\\ p_{10} & p_{11}\end{bmatrix}=\begin{bmatrix}0.9 & 0.1\\ 0.2 & 0.8\end{bmatrix}$$

如图 8.5-2 所示。

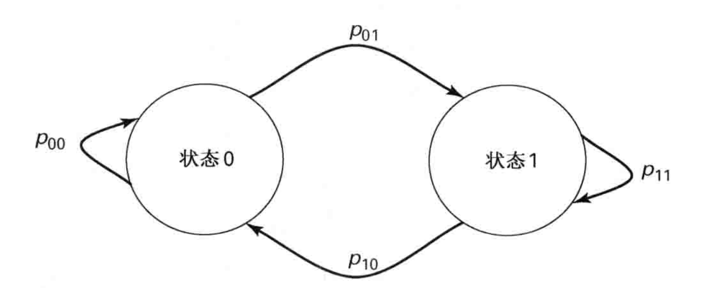


图 8.5-2　一般(非对称)二状态马尔可夫随机序列的状态转移图，状态标签 0 和 1

注意在此模型中，不要求 $p_{00}=p_{11}$，而稳态概率，如果存在的话，由以下方程的解确定

$$\boldsymbol{p}[n+1]=\boldsymbol{p}[n]\boldsymbol{P} \tag{8.5-5}$$

其中令 $n\to\infty$。记概率为 $p_0[\infty]$ 和 $p_1[\infty]$，利用 $p_0[\infty]+p_1[\infty]=1$，得到

$$p_0[\infty]=\frac{1-p_{11}}{2-p_{00}-p_{11}}$$

$$p_1[\infty]=\frac{1-p_{00}}{2-p_{00}-p_{11}}$$

由此，对于例 8.5-4 中的 $\boldsymbol{P}$ 矩阵，计算得到 $p_0[\infty]=\dfrac{2}{3}$ 以及 $p_1[\infty]=\dfrac{1}{3}$。

本例中 AMC 的稳态自相关函数可以由马尔可夫状态概率计算得到。例如，假设渐近平稳

$$\begin{aligned}R_{XX}[m] &\approx P\{X[k]=1, X[m+k]=1\}, \text{足够大的}k\\&= P\{X[k]=1\}P\{X[m+k]=1|X[k]=1\}\\&= p_1[\infty]\ P\{X[m]=1|X[0]=1\}\end{aligned} \tag{8.5-6}$$

其中最后一个因子是从状态1到状态1的 m 步转移概率。当初始条件 $\boldsymbol{p}[0]=[0, 1]$，可以递归地由式(8.5-5)计算得到，所需要的计算也可以用下例中的格状图说明。

例 8.5-6 (马尔可夫链的格状图)再次考虑例 8.1-15，为目前所知道的马尔可夫链引入了状态转移图。另一种展示到达指定状态的可能路径及其概率的有用图形是格状图，名字源于常见的围绕某些农田的木质篱笆。图 8.5-3 给出了二状态的情况，状态标签为 0 和 1，假设是对称的，即 $p_{ij}=p_{ji}$。可以发现这个格状图是例 8.1-4 中更一般的树状图的折叠。之所以可以将树折叠成格，是因为马尔可夫条件是基于条件概率的，条件概率充当了分支的标签。

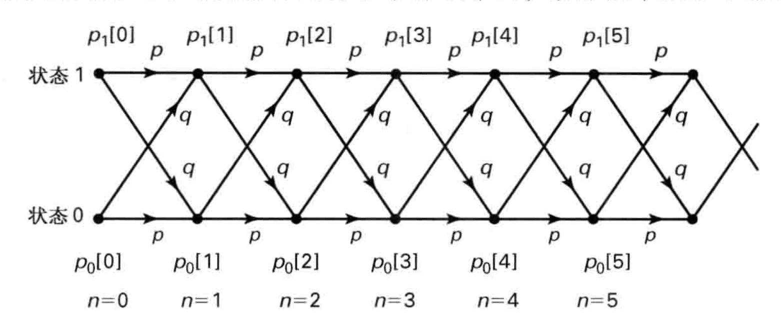


图 8.5-3 状态标签为 0 和 1 的二状态对称马尔可夫链。$p_i[n]$ 表示时刻 n 处于状态 i 的概率

每个节点代表给定时刻的状态。节点的值(标签)是其在时刻 n 的概率。连线(有向分支)表示可能的状态转移，标注了相应的状态转移概率。通过格状图的路径表示可能的多步状态转移，其概率为路径上每个分支转移概率的乘积。

如果已经知道在时刻 $n=0$ 链的状态为1，则改进的格状图简化为图 8.5-4，其中在状态1的节点上标注了 $P_n \triangleq P\{X[n]=1|X[0]=1\}$，则可以利用此格状图计算式(8.5-6)所需的概率 $P\{X[n]=1|X[0]=1\}$。前面的几个 P_n 很容易计算出来，如 $P_1=p$，$P_2=p^2+q^2$，$P_3=p^3+3pq^2$ 等。对于 $p_0[\infty]=p_1[\infty]=\frac{1}{2}$，且 $p=0.8$ 的情况，渐近平稳自相关(ASA)函数 $R_{XX}[m]$ 为 $R_{XX}[0]=0.5$，$R_{XX}[\pm1]=0.4$，$R_{XX}[\pm2]=0.34$，$R_{XX}[\pm3]=0.304$，以此类推[①]。

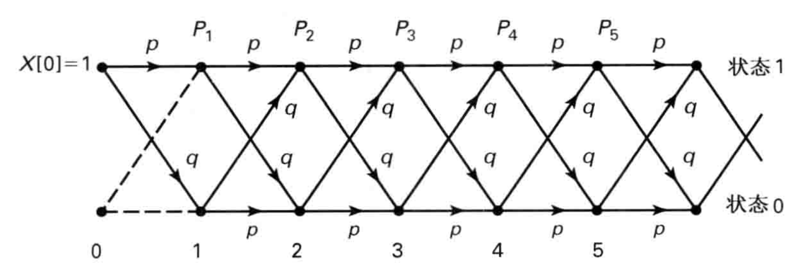


图 8.5-4 基于条件 $X[0]=1$ 的格状图

① ASA 是通过 $R_{XX}[m]=E\{X[k+m]X[k]\}$，其中 $k\to\infty$ 计算的。对于 0 和 1 两级，$R_{XX}[m]=P\{X[m+k]=1|X[k]=1\}\times0.5$，则显然 $R_{XX}[0]=1\times0.5=0.5$，$R_{XX}[1]=0.8\times0.5=0.4$，$R_{XX}[2]=[(0.8)^2+(0.2)^2]\times0.5=0.34$ 等。

除了一些简单的情况，格状图一般会显示存在许多可能的路径到达特定节点，即给定时间的给定状态。这就产生了哪条路径最有可能(可能性最大)产生要求的多步转移的路径的问题。在上例中，若取 $p>q$，这只是找到具有最多 p 的路径问题。但是一般情况下，找到最可能的路径是一件非常耗时的任务，而且如果采用随机测试的方法，则很快就会超越多数计算机的处理能力。针对这个问题开展了许多研究，源于其广泛的工程应用需求，其中就包括机器语音识别。

例 8.5-7 (缓冲占用)考虑马尔可夫链作为具有 $M+1$ 个状态的通信信道缓冲区的模型，用标签 0 到 M 表示缓冲的占用数。换句话说，状态标签是 M 字节容量的缓冲中目前存储的字节数量。假设状态转移只能发生在相邻状态之间；也就是在每个时间单元，占用量最多变化一个字节，那么其状态转移图如图 8.5-5 所示。

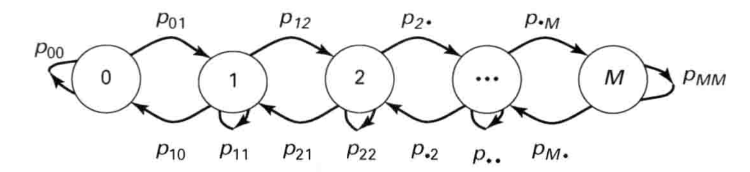


图 8.5-5 $M+1$ 状态通信缓冲区的马尔可夫链模型

如果在例 8.5-7 中令 M 趋于无穷，则得到所谓的一般生灭链，首先被用来对人口随时间变化的规模建模。在每个时间单元，存在最多一个出生和最多一个死亡。

问题求解 考虑一个具有如下状态转移矩阵的二状态马尔可夫链

$$\boldsymbol{P}=\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

可以写出 $\boldsymbol{p}[n]$ 与 $\boldsymbol{p}[n+1]$ 关系的方程如下：

$$[p_0[n+1],\ \ p_1[n+1]]=[p_0[n],\ \ p_1[n]]\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \tag{8.5-7}$$

这个向量方程等价于两个必须同时求解的标量差分方程，也就是两个同步差分方程。下面尝试以下形式的解：

$$p_0[n]=C_0z^n, \qquad p_1[n]=C_1z^n$$

将此尝试解代入式(8.5-7)，消除公共项 z^n，得到

$$C_0z=C_0p_{00}+C_1p_{10}$$

$$C_1z=C_0p_{01}+C_1p_{11}$$

由此推出下面的必要条件：

$$C_1=C_0\left(\frac{z-p_{00}}{p_{10}}\right)=C_0\left(\frac{p_{01}}{z-p_{11}}\right)$$

这确定了常数 C_0 和 C_1 之间的约束关系以及对 z 的必要条件，后者称为特征方程

$$(z-p_{00})(z-p_{11})-p_{10}p_{01}=0$$

实际上利用行列式函数，特征方程(CE)可以写成

$$\det(z\boldsymbol{I}-\boldsymbol{P})=0$$

求解两状态方程，得到两个解 z_1 和 z_2，其中之一必须等于 1(看出来了吗？注意 $1-p_{00}=p_{01}$。)由此得到的解可以写成

$$p_0[n]=C_0z_i^n, p_1[n]=C_0\left(\frac{z_i-p_{00}}{p_{10}}\right)z_i^n, i=1,2$$

由于向量差分方程是线性的，可以给不同的 z_i 值增加两个相应的解，得到一般解，写成行向量形式如下：

$$\boldsymbol{p}[n]=A_1\left[1,\frac{z_1-p_{00}}{p_{10}}\right]z_1^n+A_2\left[1,\frac{z_2-p_{00}}{p_{10}}\right]z_2^n$$

其中对两个线性独立解引入了两个新的常数 A_1 和 A_2。这两个常数必须由以下条件求解，初始概率矩阵 $\boldsymbol{p}[0]$ 和时刻 n 的概率行向量的必要条件，即对所有 $n\geqslant 0$，有 $\sum_{i=0}^{1}p_i[n]=1$。

例 8.5-8 （完全解）令

$$\boldsymbol{P}=\begin{bmatrix}0.9 & 0.1\\0.2 & 0.8\end{bmatrix},\qquad \boldsymbol{p}[0]=[1/2,\ 1/2]$$

并求完全解，包括初始转移和 $\boldsymbol{p}[n]$ 的稳态值。

第一步是求出 $\boldsymbol{P}$ 的特征值，也就是特征方程(CE)的根

$$\det(z\boldsymbol{I}-\boldsymbol{P})=\det\begin{bmatrix}z-0.9 & -0.1\\-0.2 & z-0.8\end{bmatrix}=z^2-1.7z+0.7=0$$

此方程的根为 $z_1=0.7$ 和 $z_2=1.0$。因此可以写为

$$\boldsymbol{p}[n]=C_1[1,-1]\,0.7^n+C_2[1,0.5]\,1^n$$

由稳态要求 $\boldsymbol{p}$ 的元素求和为 1.0，可以得到 $C_2=\frac{2}{3}$。因此可以进一步写为

$$\boldsymbol{p}[n]=C_1[1,-1]\,0.7^n+\left[\tfrac{2}{3},\tfrac{1}{3}\right]$$

最后利用指定的初始条件 $\boldsymbol{p}[0]=[1/2,1/2]$ 得到 $C_1=-\frac{1}{6}$ 及

$$\boldsymbol{p}[n]=\left[-\tfrac{1}{6},\tfrac{1}{6}\right]0.7^n+\left[\tfrac{2}{3},\tfrac{1}{3}\right],\quad \text{或者以标量形式}$$

$$\begin{aligned}p_0[n]&=-\tfrac{1}{6}\,0.7^n+\tfrac{2}{3}\\p_1[n]&=\tfrac{1}{6}\,0.7^n+\tfrac{1}{3}\end{aligned},\qquad n\geqslant 0$$

可以看到稳态概率存在，且 $p_0[\infty]=\frac{2}{3}$ 及 $p_1[\infty]=\frac{1}{3}$。下面的例子将说明这种稳态概率并不总是存在的。

例 8.5-9 （乒乓序列）考虑状态转移概率矩阵为 $\boldsymbol{P}=\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$ 的二状态马尔可夫链。其特征方程为

$$\det(z\boldsymbol{I}-\boldsymbol{P})=\det\begin{bmatrix}z & -1\\-1 & z\end{bmatrix}=z^2-1=0$$

有两个根 $z_{1,2}=\pm 1$。因此这种情况下没有稳定的状态，即使 $\boldsymbol{P}$ 的特征值之一为 1。实际上，可以从状态转移图直接看到，此随机序列永远在每个连续的时间周期在状态 0 和 1 之间循环往复。其相位可以由初始概率向量 $\boldsymbol{p}[0]=[1,0]$ 不可变确定。

我们不能总是假设稳定状态存在，注意此例子的状态转移概率退化到达和离开状态的概

率要么是0，要么是1。另一个关于稳态存在性的问题是所谓的捕获状态。这是一种只有到达，没有离开的状态。在大多数通信和信号处理所感兴趣的场合，稳态都是存在的，不管链的初始状态如何。

8.6　向量随机序列和状态方程

目前为止接触过的标量随机序列的概念可以扩展到向量随机序列。此概念用来导出噪声中信号的线性估计(卡尔曼滤波)(参见第11章)。也被用在传感器阵列的建模中，例如地震、声学、雷达等。本节介绍随机向量的差分方程和向量马尔可夫随机序列的概念。有趣的是，高阶马尔可夫-p标量随机序列可以看成一阶向量马尔可夫序列。

定义8.6-1　一个向量随机序列是从概率样本空间Ω，对应于概率空间$(\Omega, \mathscr{F}, P)$，到复数向量序列空间的映射。

因此对于每个$\zeta\in\Omega$和固定的时刻n，得到一个向量$\boldsymbol{X}(n, \zeta)$。向量随机序列通常写为$\boldsymbol{X}[n]$，省略实验结果ζ。

例如随机向量序列$\boldsymbol{X}[n]$的一阶累积分布函数可以记为

$$F_{\boldsymbol{X}}(\boldsymbol{x};n) \triangleq P\{\boldsymbol{X}[n] \leqslant \boldsymbol{x}\}$$

其中$\{\boldsymbol{X}[n]\leqslant\boldsymbol{x}\}$意味着每个元素均满足此不等式，即$\{X_1[n]\leqslant x_1, X_2[n]\leqslant x_2, \cdots, X_N[n]\leqslant x_N\}$。二阶及更高阶概率也可以相应地进行描述。向量随机序列由其所有一阶和高阶累积分布函数(或概率密度函数、概率质量函数)统计地确定。

下面的例子对向量随机序列输入以下具有N维系数矩阵$\boldsymbol{A}$和$\boldsymbol{B}$的向量LCCDE进行相关性分析

$$\boldsymbol{y}[n] = \boldsymbol{A}\boldsymbol{y}[n-1] + \boldsymbol{B}\boldsymbol{x}[n]$$

对于此向量，当$\boldsymbol{A}$的特征值的绝对值小于1，可以假定是BIBO稳定的。

例8.6-1　(向量LCCDE)对于向量的情况，标量的一阶LCCDE模型，由列向量随机序列$\boldsymbol{X}[n]$激励，成为

$$\boldsymbol{Y}[n] = \boldsymbol{A}\boldsymbol{Y}[n-1] + \boldsymbol{B}\boldsymbol{X}[n] \tag{8.6-1}$$

这是在样本向量序列中的一阶向量差分方程。向量冲激响应为列向量序列

$$\boldsymbol{h}[n] = \boldsymbol{A}^n\boldsymbol{B}u[n]$$

序列$\boldsymbol{X}[n]$的零初始条件响应为

$$\begin{aligned}\boldsymbol{Y}[n] &= \sum_{k=0}^{n}\boldsymbol{A}^{n-k}\boldsymbol{B}\boldsymbol{X}[k] \\ &\triangleq \boldsymbol{h}[n] * \boldsymbol{X}[n]\end{aligned}$$

其矩阵系统函数为

$$\boldsymbol{H}(z) = (\boldsymbol{I} - \boldsymbol{A}z^{-1})^{-1}\boldsymbol{B}$$

可以很容易地验证。输入WSS随机向量序列及其输出WSS随机序列之间的WSS互相关矩阵为$\boldsymbol{R}_{\boldsymbol{YX}}[m]\triangleq E\{\boldsymbol{Y}[n+m]\boldsymbol{X}^{\dagger}[n]\}$[其中符号“†”表示埃尔米特(或共轭)转换]和$\boldsymbol{R}_{\boldsymbol{YX}}[m]\triangleq E\{\boldsymbol{X}[n+m]\boldsymbol{Y}^{\dagger}[n]\}$成为

$$\boldsymbol{R}_{\boldsymbol{YX}}[m]=\boldsymbol{h}[m]*\boldsymbol{R}_{\boldsymbol{XX}}[m]$$

$$\boldsymbol{R}_{\boldsymbol{XY}}[m]=\boldsymbol{R}_{\boldsymbol{XX}}[m]*\boldsymbol{h}^{\dagger}[-m]$$

顺便地，我们注意到对于因果的 $\boldsymbol{h}$，可从上面的向量 LCCDE 递归解得到，当输入 $\boldsymbol{X}=\boldsymbol{W}$ 假设为一白噪声向量序列，输出 $\boldsymbol{Y}[n]$ 与输入 $\boldsymbol{X}[n]$ 将来的取值无关。

输出的相关矩阵为

$$\boldsymbol{R}_{\boldsymbol{YY}}[m]=\boldsymbol{h}[m]*\boldsymbol{R}_{\boldsymbol{XX}}[m]*\boldsymbol{h}^{\dagger}[-m]$$

进行傅里叶变换，输出 psd 矩阵为

$$\boldsymbol{S}_{\boldsymbol{YY}}(\omega)=\boldsymbol{H}(\omega)\boldsymbol{S}_{\boldsymbol{XX}}(\omega)\boldsymbol{H}^{\dagger}(\omega)$$

式(8.6-1)对于任意 $n\geqslant n_0$ 的完全解可以用 n_0 时刻明确的初始条件 $\boldsymbol{Y}[n_0]$ 写为

$$\boldsymbol{Y}[n]=\boldsymbol{A}^{n-n_0}\boldsymbol{Y}[n_0]+\sum_{k=n_0}^{n}\boldsymbol{h}[n-k]\boldsymbol{X}[k],\qquad n\geqslant n_0$$

当 $n_0\to-\infty$，对稳定的系统矩阵 $\boldsymbol{A}$，这变成卷积和

$$\boldsymbol{Y}[n]=\boldsymbol{h}[n]*\boldsymbol{X}[n],\qquad -\infty<n<+\infty$$

定义 8.6-2 向量随机序列 $\boldsymbol{Y}[n]$ 为向量马尔可夫序列，如果对所有 $K>0$ 以及对所有 $n_K>n_{K-1}>\cdots>n_1$，下式

$$P\{\boldsymbol{Y}[n_K]\leqslant\boldsymbol{y}_K|\boldsymbol{y}[n_{K-1}],\cdots,\boldsymbol{y}[n_1]\}=P\{\boldsymbol{Y}[n_K]\leqslant\boldsymbol{y}_K|\boldsymbol{y}[n_{K-1}]\}$$

对所有向量 $\boldsymbol{y}_K$，以及所有条件向量 $\boldsymbol{y}[n_{K-1}],\cdots,\boldsymbol{y}[n_1]$ 都成立(参见定义 8.5-2 的马尔可夫-p 性质)。

下面介绍向量随机序列的一个定理。

定理 8.6-1 对于状态方程

$$\boldsymbol{X}[n]=\boldsymbol{A}\boldsymbol{X}[n-1]+\boldsymbol{B}\boldsymbol{W}[n],\qquad n>0,\qquad \boldsymbol{X}[0]=\boldsymbol{0}$$

令输入 $\boldsymbol{W}[n]$ 为白高斯随机序列。则对于 $n>0$ 输出 $\boldsymbol{X}[n]$ 是向量马尔可夫随机序列。

定理的证明留做练习。

例 8.6-2 (标量马尔可夫-p 和向量马尔可夫之间的关系)令 $X[n]$ 为马尔可夫-p 随机序列，满足 p 阶差分方程

$$X[n]=a_1X[n-1]+\cdots+a_pX[n-p]+bW[n]$$

定义 p 维向量随机序列 $\boldsymbol{X}[n]=[X[n],\cdots,X[n-p+1]]^{\mathrm{T}}$，以及系数矩阵

$$\boldsymbol{A}=\begin{bmatrix} a_1 & a_2 & \cdots & \cdots & a_p \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

则有

$$\boldsymbol{X}[n]=\boldsymbol{A}\boldsymbol{X}[n-1]+\boldsymbol{b}\boldsymbol{W}[n]$$

由此 $\boldsymbol{X}[n]$ 是 $\boldsymbol{b}=[b,0,\cdots,0]^{\mathrm{T}}$ 的向量马尔可夫随机序列。这种标量方程的向量转化称为状态变量表示[8-7]。

8.7　随机序列的收敛

当序列趋于无穷大，例如随着时间变为无限大，某些非平稳随机序列可能收敛到一个极限值。这种渐近行为在概率论中的无限伯努利序列的收敛性已经见证过，相应的理论称为大数定理。此外，在学习随机过程的收敛时，有时会要求一个序列越来越逼近随机系统给定时间的输出，例如 t_0 的输出 $Y_n(t_0)$。下标 n 定义了一个随机序列，在某种意义上将收敛到真实的输出。本节将研究几种随机序列，也就是随机变量序列的收敛。

首先回顾确定序列收敛的概念。令 x_n 为一复数(或实数)序列，则收敛定义如下所述。

定义 8.7-1　一复数(或实数)序列 x_n 收敛到某个复数(或实数)x，如果给定任意 $\varepsilon>0$，存在一个整数 n_0 满足当 $n>n_0$ 时，有

$$|x_n - x| < \varepsilon$$

注意在此定义中 n_0 可能依赖于 ε 的值；也就是当 ε 变得越来越小，则 n_0 很可能变得越来越大。有时候这种依赖性被形式化到定义中，用 $n_0(\varepsilon)$ 代替 n_0。这经常写为

$$\lim_{n\to\infty} x_n = x \quad \text{或当} \quad x_n \to x \quad \text{当} \quad n\to\infty$$

这种定义的一个具体问题是必须知道极限 x 才能验证其收敛性。对于简单的情况往往可以猜测出极限的值并用定义验证极限确实存在。幸运的是，对于大多复杂的情况，存在一种替代的柯西收敛准则，下面不加证明地给出。

定理 8.7-1　(柯西准则[8-8])一个复数(或实数)序列 x_n 收敛到某个极限，当且仅当(iff)

$$|x_n - x_m| \to 0 \quad \text{随着} \quad n \text{ 和 } m\to\infty$$

它对复数(实数)序列有效的原因是所有复数(或实数)集合是完备的，意味着它包含所有极限点。例如，集合$\{0<x<1\}=(0,1)$不是完备的，而集合$\{0\leqslant x\leqslant 1\}=[0,1]$是完备的，因为这些集合中趋向于 0 或 1 的序列 x_n 在$[0,1]$中有极限，而在$(0,1)$中极限点不存在。实际上，所有复数(或实数)的集合是完备的，而且 n 维线性向量空间在实数域以及复数域上都是完备的。因此柯西收敛准则也适用于这些情况。关于数值收敛的更多信息可以参阅参考文献[8-8]。

函数的收敛定义利用了数值序列收敛的定义。我们称函数序列 $f_n(x)$ 收敛到函数 $f(x)$，如果对每个 x 相应的数值序列都收敛。在下面的定义中给出了更加正式的定义。

定义 8.7-2　函数序列 $f_n(x)$(逐点)收敛到函数 $f(x)$，如果对每个 x_0，复数序列 $f_n(x_0)$ 收敛到 $f(x_0)$。

如果考虑的函数集合是完备的，则柯西收敛准则可以应用于函数的逐点收敛。连续函数的集合不是完备的，因为连续函数的序列可能收敛到非连续的函数[参见例 8.7-1 的选项(d)]。但是有限函数的集合是完备的[8-8]。

下面是收敛数值和函数序列的一些例子。验证其结果作为练习留给读者。

例 8.7-1　(一些收敛序列)

(a) $x_n = (1-1/n)a + (1/n)b \to a$　当 $n\to\infty$。

(b) $x_n = \sin(\omega + \mathrm{e}^{-n}) \to \sin\omega$　当 $n\to\infty$。

(c) $f_n(x) = \sin[(\omega + 1/n)x] \to \sin(\omega x)$，　当 $n\to\infty$，对于任意(固定)x。

(d) $f_n(x) = \left\{\begin{matrix} \mathrm{e}^{-n^2 x}, & x>0 \\ 1, & x\leqslant 0 \end{matrix}\right\} \to u(-x)$，　当 $n\to\infty$，对于任意(固定)x。

读者需要注意对于(c)和(d)的函数收敛，当取极限的时候，变量 x 是保持不变的。对每个 x 重复求极限以求出极限函数。

因为随机变量是一个函数，所以随机变量的序列(也称为随机序列)就是函数的序列。因此，可以定义第一类，也是最强的随机变量收敛类型。

定义 8.7-3 (确定收敛)随机序列 $X[n]$ 确定收敛到随机变量 X，如果函数的序列 $X[n,\zeta]$ 收敛到函数 $X(\zeta)$，当 $n\to\infty$ 且对所有实验结果 $\zeta\in\Omega$。

提醒一下，函数 $X(\zeta)$ 不是任意的。它们是随机变量，因此满足以下条件，对所有 x 集合 $\{\zeta: X(\zeta)\leqslant x\}\subset\mathscr{F}$，也就是，此集合是对所有 x 值的事件。这对于计算概率而言实际上是必需的，因为概率测度 P 只对事件有定义。这样的函数 X 通常被称为测度函数，而且在实数分析课程中验证了测度函数的空间是完备的[8-1]。若有一测度函数(随机变量)的柯西序列，那么可以证明极限函数存在且可测(一个随机变量)。因此柯西收敛准则同样适用于随机变量。

大多数时候我们对 Ω 中概率为零的随机变量集合不感兴趣，因为考虑到这些事件永远不会出现。在这种情况下，可以弱化确定收敛的定义，得到仍然非常强的几乎确定收敛的定义。

定义 8.7-4 (几乎确定收敛)随机序列 $X[n]$ 几乎确定收敛到随机变量 X，如果函数的序列 $X[n,\zeta]$ 除了可能在概率为零的集合上，对所有实验结果 $\zeta\in\Omega$ 收敛。

这是通常在概率论中最强的收敛类型，也称为概率 1 收敛。有时候记为

$$P\left\{\lim_{n\to\infty} X[n,\zeta]=X(\zeta)\right\}=1$$

简单来说就是存在集合 A 满足 $P[A]=1$ 且对所有 $\zeta\in A$，$X[n]$ 收敛到 X。特别地 $A\triangleq\{\zeta:\lim_{n\to\infty}X[n,\zeta]=X(\zeta)\}$。这里集合 A^c 是定义中提到的概率零集合。简记为

$$X[n]\to X \text{ a.s.} \quad \text{和} \quad X[n]\to X \quad \text{pr.1}$$

其中缩写“a. s”表示几乎确定，而“pr. 1”表示概率 1。

概率 1 收敛的一个例子是强大数定理，将在下一节证明。接下来讨论可能收敛的随机序列的三个例子。

例 8.7-2 (随机序列的收敛)下面三个随机序列，都假设概率空间$(\Omega,\mathscr{F},P)$有样本空间 $\Omega=[0,1]$。$\mathscr{F}$是 Ω 的博雷尔子集，概率测度 P 为勒贝格测度，在实区间$(a,b]$就是其长度 l，即

$$l(a,b]\triangleq b-a, \qquad b\geqslant a$$

(a) $X[n,\zeta]=n\zeta$
(b) $X[n,\zeta]=\sin(n\zeta)$
(c) $X[n,\zeta]=\exp[-n^2(\zeta-\frac{1}{n})]$

显然(a)中的序列对任意 $\zeta\neq 0$ 发散到 $+\infty$。因此这个随机序列不收敛。(b)中的序列不发散，但是除了 $\zeta=0$，其在 -1 和 $+1$ 之间振荡，因此也不收敛。

考虑(c)中的随机序列，图 8.7-1 显示此序列收敛如下：

$$\lim_{n\to\infty} X[n,\zeta]=\begin{cases}\infty, & \zeta=0\\ 0, & \zeta>0\end{cases}$$

因此可以说，随机序列以概率1收敛到(退化的)随机变量 $X=0$。简单取 $A=(0,1]$ 且注意到 $P[A]=1$ 及对 A 中每个 ζ，当 n 足够大，$X[n,\zeta]\approx 0$。记为 $X[n]\to 0$，$X[n]\to 0$ 但是 $X[n]$ 显然不是确定收敛到零。

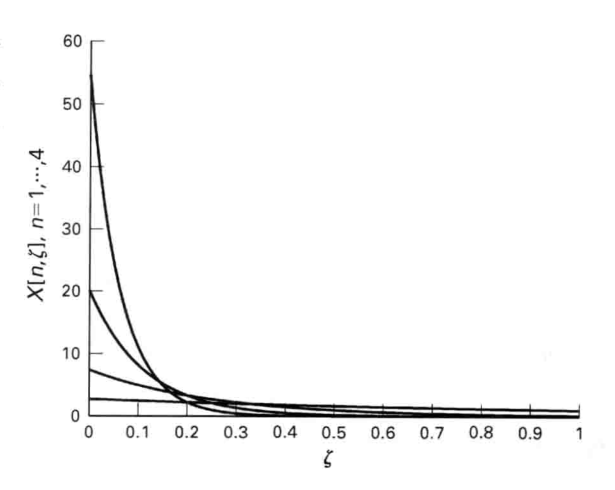


图 8.7-1 例 8.7-2(c)中序列的图形，$X[n,\zeta]$ 与 ζ，$\Omega=[0,1]$，$n=1,\cdots,4$

目前我们已经讨论了函数序列和随机序列的逐点收敛。这类似于考虑具有以下范数的有界函数空间 $\mathscr{B}$:

$$||f||_\infty \triangleq \sup_x |f(x)|^\dagger$$

当在函数空间 $\mathscr{B}$ 中表示 $f_n\to f$ 时，意思是 $\|f_n-f\|_\infty=\sup_x|f_n(x)-f(x)|\to 0$，提供逐点收敛的定义。连续有界函数的空间记为 L_∞，它是完备的[8-1]。

另一种具有很强实际意义的函数空间类型采用能量范数[参见式(4.4-6)]

$$||f||_2 \triangleq \left(\int_{-\infty}^{+\infty} |f(x)|^2 \mathrm{d}x\right)^{1/2}$$

具有有限能量范数的可积(可测)函数空间记为 L^2。当称一个函数序列在 L^2 中收敛，即 $\|f_n-f\|_2\to 0$，意味着

$$\left(\int_{-\infty}^{+\infty} |f_n(x)-f(x)|^2 \mathrm{d}x\right)^{1/2} \longrightarrow 0 \quad 当 \quad n\longrightarrow\infty$$

可积函数的空间也是完备的[8-1]。相应的随机序列的概念称为均方收敛。

定义 8.7-5 (均方收敛)称随机序列 $X[n]$ 均方收敛到随机变量 X，如果 $E\{|X[n]-X|^2\}\to 0$，$n\to\infty$。

这种收敛仅依赖于随机变量的二阶属性，因此往往比 a.s 收敛容易计算。均方收敛的第二个优点是其和物理意义上的功率概念密切相关。如果 $X[n]$ 在均方意义上收敛到 X，那么可以预计，当 n 比较大时，其误差的方差 $\varepsilon[n]\triangleq X[n]-X$ 比较小。如果回头再看看例 8.7-2(c)，可以发现此随机序列在均方意义上并不收敛，因此就这里所定义的误差的方差或功率而言，不可能如期望的那样小。为了看清这一点，假定均方收敛到零(因 $X[n]\to 0$)

$$\begin{aligned}
E\{|X[n]-0|^2\} &= E\{X[n]^2\}\\
&= \int_0^1 \exp(-2n^2\zeta)\exp 2n\mathrm{d}\zeta\\
&= \exp(2n)\int_0^1 \exp(-2n^2\zeta)\mathrm{d}\zeta\\
&= \exp(2n)\left[\frac{1-\exp(-2n^2)}{2n^2}\right] \to\infty \quad 随着\ n\to\infty
\end{aligned}$$

因此从均方意义上 $X[n]$ 并不收敛到0。

还有一种将要考虑的收敛类型被称为依概率收敛。它比概率1收敛弱，也比均方收敛弱。这种类型的收敛出现在下一节将讨论的弱大数定理中。其定义如下所述。

定义 8.7-6 (依概率收敛)给定随机序列 $X[n]$ 及极限随机变量 X，称 $X[n]$ 依概率收敛到 X，如果对任意 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\left[|X[n]-X|>\varepsilon\right]=0$$

有时记为 $X[n]\to X(p)$，其中(p)表示收敛的类型。依概率收敛有时也称为 p 收敛。

可以利用切比雪夫不等式(参见定理4.4-1)，对 $\varepsilon>0$，$P[|Y|>\varepsilon]\leqslant E[|Y|^2]/\varepsilon^2$ 来验证均方收敛隐含依概率收敛。例如，令 $Y\triangleq X[n]-X$，则前面的不等式变为

$$P[|X[n]-X|>\varepsilon]\leqslant E\left[|X[n]-X|^2\right]/\varepsilon^2$$

现在均方收敛意味着当 $n\to\infty$ 时右端趋向于零，对任何确定的 $\varepsilon>0$，这意味着左边必须趋向于零，恰好是依概率收敛的定义。由此证明了下面的结论。

定理 8.7-2 随机序列均方收敛则必依概率收敛。

概率1收敛与依概率收敛的关系更加微妙。注意到前者讨论极限的概率，而后者讨论概率的极限，就可以发现它们之间的主要区别。此外，注意到 a.s 收敛关注整个样本序列的收敛，而 p 收敛只关注单个 n 的随机变量的收敛，可以获得更深入的认识。也就是说，a.s 收敛关注无限数量时间点的联合事件，而 p 收敛关注的是时刻 n 的事件，无论它有多大。可以证明下面的定理。

定理 8.7-3 概率1收敛则依概率收敛

证明 (参考了 Gnedenko[8-9])令 $X[n]\to X$ 且定义集合 A

$$A\triangleq\bigcap_{k=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{m=1}^{\infty}\{\zeta: |X[n+m,\zeta]-X(\zeta)|<1/k\}$$

则必有 $P[A]=1$。为说明这一点，注意 A 是满足从某个 n 开始的，对所有后续 n 都有 $|X[n,\zeta]-X(\zeta)|<1/k$ 的 ζ 的集合，而且此式对所有 $k>0$ 都成立。因此 A 恰为使 $X[n,\zeta]$ 收敛的 ζ 的集合，所以 $P[A]$ 必为1。最终，当 n 足够大，而且 $1/k$ 足够小，可以得到 $|X[n,\zeta]-X[\zeta]|<\varepsilon$，对所有更大的 n，误差保持为这么小。因此

$$P\left[\bigcup_{n=1}^{\infty}\bigcap_{m=1}^{\infty}\{|X[n+m]-X|<\varepsilon\}\right]=1,\ \text{对于所有}\ \varepsilon>0$$

根据概率的连续性，这意味着

$$\lim_{n\to\infty}P\left[\bigcap_{m=1}^{\infty}\{|X[n+m]-X|<\varepsilon\}\right]=1,\ \text{对于所有}\ \varepsilon>0$$

这接下来就隐含了极大弱化的结果

$$\lim_{n\to\infty}P[|X[n+m]-X|<\varepsilon]=1,\ \text{对于所有}\ \varepsilon>0 \tag{8.7-1}$$

这等价于 p 收敛的定义。

由于前面证明中大幅度弱化了 a.s 的条件，即扩大了集合 A 的范围，可以看出 p 收敛并不隐含 a.s 收敛。特别注意到即使对于所有 $n+m>n$ 没有一个样本序列接近 X，式(8.7-1)仍有可能是正确的。这实际上是这两种收敛之间最关键的差别。

例 8.7-3　(一个收敛的随机序列?)定义一个随机脉冲序列 $X[n]$，$n\geq 0$ 如下：令 $X[0]=1$。然后对后面的两个点，按照相同的可能性设其中一个点 $X[n]$ 为 1，另一个为 0。接着对后面的三个点，按照相同的可能性设其中一个点 $X[n]$ 为 1，另两个为 0。继续这一过程，对后面的四个点，按照相同的可能性设其中一个点 $X[n]$ 为 1，另三个为 0，以此类推。一个样本函数如图 8.7-2 所示。



图 8.7-2　依概率收敛但不是概率 1 收敛的序列

显然此随机序列在某种意义上，当 $n\to\infty$ 时缓慢收敛到零。实际上通过简单的计算可以证明 p 收敛，以及均方收敛，由于随着 $n\to\infty$，脉冲之间的距离逐步加大。实际上在 $n\approx\frac{1}{2}l^2$ 时，1(脉冲)出现的概率仅为 $1/l$。但是，得不到 a.s 收敛，因为每个样本序列中，在 n 轴任意远的地方都有 1。因此没有样本序列收敛到零。

最后一种考虑的收敛类型根本不是随机变量的收敛！而是分布函数的收敛。

定义 8.7-7　具有累积分布函数 $F_n(x)$ 的随机序列 $X[n]$ 收敛到具有累积分布函数 $F(x)$ 的随机变量 X，如果

$$\lim_{n\to\infty} F_n(x) = F(x)$$

对所有 x，F 是连续的。

注意在此定义中，说的不是随机变量自身，而是它们的累积分布函数。分布收敛只意味着当 n 变大，累积分布函数收敛或者变得类似。例如，序列 $X[n]$ 和变量 X 可能联合独立，即使 $X[n]$ 分布收敛到 X。这与前面四种收敛完全不同，对于那些收敛，当 n 变大，随机变量 $X[n]$ 和 X 变得非常相关，因为两者之间某种类型的“误差”趋近于零。分布收敛是出现在中心极限定理(参见 4.7 节)中的收敛类型。这 5 种收敛之间的关系以图形的方式在图 8.7-3 中给出，其中利用了 p 收敛隐含依概率收敛，下面会说明这一点。注意，即使确定收敛也不一定意味着均方收敛。这是因为随机变量极限平方的积分，相对于概率测度，可能发散。

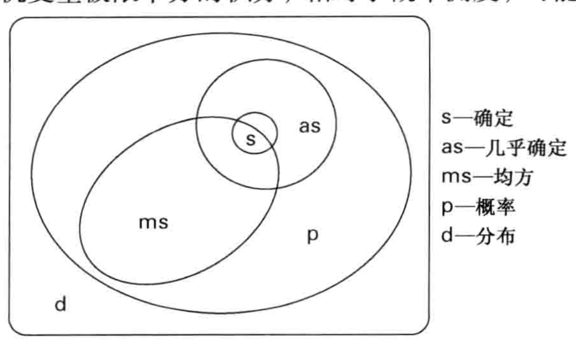


图 8.7-3　随机序列各种可能收敛之间关系的维恩图

为了说明 p 收敛隐含分布收敛，假设极限随机变量 X 是连续的，因此其具有概率密度函数。首先考虑条件分布函数

$$F_{X[n]|X}(y|x) = P\{X[n] \leqslant y | X = x\}$$

由 p 收敛的定义，显然

$$F_{X[n]|X}(y|x) \to \begin{cases} 1, y > x \\ 0, y < x \end{cases}, \qquad 当\ n \to \infty$$

因此

$$F_{X[n]|X}(y|x) \to u(y-x)$$

除非在点 $y = x$。因此有

$$\begin{aligned} F_{X[n]}(y) = P\{X[n] \leqslant y\} &= \int_{-\infty}^{+\infty} F_{X[n]|X}(y|x) f_X(x) \mathrm{d}x \\ &\to \int_{-\infty}^{+\infty} u(y-x) f_X(x) \mathrm{d}x \\ &= \int_{-\infty}^{y} f_X(x) \mathrm{d}x \\ &= F_X(y) \end{aligned}$$

得证。当极限随机变量 X 不连续，那么应用的时候必须谨慎，但是对于在 $F_X(x)$ 连续的所有点 x 上，结论仍然成立(参见习题 8.52)。

8.8 大数定理

大数定理是关于随机变量均值估计序列的收敛。因此其关注随机序列收敛到常数。弱大数定理得到依概率收敛，强大数定理产生概率 1 收敛。弱大数定理的一种说法已经在例 4.4-3 中介绍过。为方便计，这里复述一遍。

定理 8.8-1 (弱大数定理)令 $X[n]$ 为定义在 $n \geqslant 1$ 上，均值为 μ_X，方差为 σ_X^2 的独立随机序列。定义另一随机序列如下：

$$\hat{\mu}_X[n] \triangleq (1/n) \sum_{k=1}^{n} X[k], \qquad n \geqslant 1$$

则 $\hat{\mu}_X[n] \to \mu_x \quad (p) \quad 当\ n \to \infty$ 。

记住，一个独立随机序列的各随机变量都相互独立。弱大数定理的另一种说法允许随机序列具有不均匀的方差。

定理 8.8-2 (弱大数定理，非均匀方差)令 $X[n]$ 为定义在 $n \geqslant 1$ 上，均值为 μ_X，方差为 $\sigma_X^2[n]$ 的独立随机序列。若

$$\sum_{n=1}^{\infty} \sigma_X^2[n]/n^2 < \infty$$

则 $\hat{\mu}_X[n] \to \mu_X \quad (p) \quad 当\ n \to \infty$ 。

这两个定理对于概率 1 收敛也成立，此时变成强大数定理。关注概率 1 收敛的强大数定

理最好利用鞅序列导出。为了介绍此概念，需要另一个非常有用的结论，称为鞅收敛定理，对于估计和检测/判决理论非常有用。

定义 8.8-1①　定义在 $n\geqslant 0$ 上的随机序列 $X[n]$ 称为鞅序列，如果其条件期望为

$$E\{X[n]|X[n-1],X[n-2],\cdots,X[0]\}=X[n-1]\quad \text{对于所有 } n\geqslant 1$$

将条件期望看成基于历史数据对序列当前值的估计，则对于鞅序列，此估计恰是最近的历史取值。如果将 $X[n]$ 作为一个赌博游戏中的资金数，那么鞅条件可以看成游戏公平性的必要条件，这实际上就是它最早被提出的方式[8-1]。

例 8.8-1　（二项式计数序列）令 $W[n]$ 为定义在 $n\geqslant 0$ 上，等概率取 ± 1 值的伯努利随机序列。令 $X[n]$ 为相应的二项式计数序列

$$X[n]\triangleq\sum_{k=0}^{n}W[k],\qquad n\geqslant 0$$

则 $X[n]$ 为鞅序列，证明如下：

$$\begin{aligned}E\{X[n]|X[n-1],\cdots,X[0]\}&=E\left\{\sum_{k=0}^{n}W[k]|X[n-1],\cdots,X[0]\right\}\\&=\sum_{k=0}^{n}E\{W[k]|X[n-1],\cdots,X[0]\}\\&=\sum_{k=0}^{n}E\{W[k]|W[n-1],\cdots,W[0]\}\\&=\sum_{k=0}^{n-1}W[k]+E\{W[n]\}\\&=X[n-1]\end{aligned}$$

第一个等式是由于 $X[n]$ 的定义。第三个等式是由于知道了前面第一个 $(n-1)X$ 相当于知道了前面第一个 $(n-1)W$。倒数第二个等式是因为 $E[W|W]=W$。最后一个等式是因为 $E\{W[n]\}=0$。

例 8.8-2　（独立增量序列）令 $X[n]$ 为定义在 $n\geqslant 0$ 上的独立增量随机序列（参见定义 8.1-4），则 $X_c[n]\triangleq X[n]-\mu_X[n]$ 为鞅序列。为证明这一点，记 $X_c[n]=(X_c[n]-X_c[n-1])+X_c[n-1]$，根据独立增量性以及 X_c 的均值为零，有

$$\begin{aligned}E\{X_c[n]|X_c[n-1],\cdots,X_c[0]\}&=E\{X_c[n]-X_c[n-1]|X_c[n-1],\cdots,X_c[0]\}\\&\quad+E\{X_c[n-1]|X_c[n-1],\cdots,X_c[0]\}\\&=E\{X_c[n]-X_c[n-1]\}+X_c[n-1]\\&=X_c[n-1]\end{aligned}$$

下面的定理显示了强大数定理与鞅序列之间的关系。强大数定理关注样本序列收敛，提供了一种在 n 点鞅序列中关于最大项的切比雪夫不等式。

① 在首次阅读的时候可以略过鞅序列的内容。

定理 8.8-3　令 $X[n]$ 为定义在 $n \geqslant 0$ 上的鞅序列，则对于每个 $\varepsilon > 0$ 和任意正的 n，有

$$P\left[\max_{0 \leqslant k \leqslant n} |X[k]| \geqslant \varepsilon\right] \leqslant E\{X^2[n]\}/\varepsilon^2$$

证明　对于 $0 \leqslant j \leqslant n$，定义互斥事件

$$A_j \triangleq \{|X[k] \geqslant \varepsilon,\ \text{在 } j \text{ 时刻的第一次}\}$$

则事件 $\{\max_{0 \leqslant k \leqslant n} |X[k]| \geqslant \varepsilon\}$ 是这些事件的联合。同时定义随机变量

$$I_j \triangleq \begin{cases} 1, & A_j \text{ 发生} \\ 0, & \text{其他} \end{cases}$$

称为事件 A_j 的指示器。则

$$E\{X^2[n]\} \geqslant \sum_{j=0}^{n} E\{X^2[n] I_j\} \tag{8.8-1}$$

因为 $\sum_{j=0}^{n} I_j \leqslant 1$，而且 $X^2[n] = (X[j] + (X[n] - X[j]))^2$，展开并将其代入式(8.8-1)，可得到

$$\begin{aligned} E\{X^2[n]\} &\geqslant \sum_{j=0}^{n} E\{X^2[j] I_j\} + 2\sum_{j=0}^{n} E\{X[j]\,(X[n] - X[j])\, I_j\} \\ &\quad + \sum_{j=0}^{n} E\{(X[n] - X[j])^2\, I_j\} \qquad (8.8\text{-}2) \\ &\geqslant \sum_{j=0}^{n} E\{X^2[j] I_j\} + 2\sum_{j=0}^{n} E\{X[j]\,(X[n] - X[j])\, I_j\} \end{aligned}$$

令 $Z_j \triangleq X[j] I_j$，将式(8.8-2)中的第 2 项写为 $E\{Z_j(X[n] - X[j])\}$ 同时注意到 Z_j 只依赖于 $X[0], \cdots, X[j]$，因此有

$$\begin{aligned} E\{Z_j\,(X[n] - X[j])\} &= E\{E[Z_j\,(X[n] - X[j])\,|X[0], \cdots, X[j]]\} \\ &= E\{Z_j E\,[X[n] - X[j]|X[0], \cdots, X[j]]\} \\ &= E\{Z_j\,(X[j] - X[j])\} \\ &= 0 \end{aligned}$$

由此式(8.8-2)变为

$$\begin{aligned} E\{X^2[n]\} &\geqslant \sum_{j=0}^{n} E\{X^2[j] I_j\} \\ &\geqslant \varepsilon^2 E\left\{\sum_{j=0}^{n} I_j\right\} \\ &= \varepsilon^2 P\left\{\bigcup_{j=0}^{n} A_j\right\} \\ &= \varepsilon^2 P\left\{\max_{0 \leqslant k \leqslant n} |X[k]| \geqslant \varepsilon\right\} \end{aligned}$$

定理 8.8-4　(鞅收敛定理)令 $X[n]$ 为定义在 $n \geqslant 0$ 上的鞅序列，满足

$$E\{X^2[n]\} \leqslant C < \infty \quad \text{对所有 } n \text{，对某个 } C$$

则

$$X[n] \to X \quad \text{(a.s.)} \quad \text{当}\, n \to \infty$$

其中 X 是极限随机变量。

证明 令 $m \geqslant 0$，定义 $Y[n] \triangleq X[n+m]-X[m]$，$n \geqslant 0$。则 $Y[n]$ 是一个鞅序列，因此根据定理8.8-3

$$P\left[\max_{0\leqslant k\leqslant n}|X[m+k]-X[m]| \geqslant \varepsilon\right] \leqslant \frac{1}{\varepsilon^2}E\left\{Y^2[n]\right\}$$

其中

$$\begin{aligned} E\{Y^2[n]\} &= E\{(X[n+m]-X[m])^2\} \\ &= E\{X^2[n+m]\} - 2E\{X[n+m]X[m]\} + E\{X^2[m]\} \end{aligned}$$

对于中间的一项，有

$$\begin{aligned} E\{X[m]X[n+m]\} &= E\{X[m]E[X[n+m]|X[m],\cdots,X[0]]\} \\ &= E\{X[m]X[m]\} \\ &= E\{X^2[m]\}，\text{由于}X\text{是鞅序列} \end{aligned}$$

所以对所有 m，$n \geqslant 0$ 都有

$$E\{Y^2[n]\} = E\{X^2[n+m]\} - E\{X^2[m]\} \geqslant 0 \quad \text{对于所有}\, m,n \geqslant 0 \tag{8.8-3}$$

因此 $E\{X^2[n]\}$ 必须是单调不减的。从上面 $C<\infty$ 可知，它还是有界的，因此必然收敛到某个极限值。因为其有极限，则根据式(8.8-3)，当 $m,n\to\infty$，$E\{Y^2[n]\}\to 0$。所以

$$\lim_{m\to\infty} P\left[\max_{k\geqslant 0}|X[m+k]-X[m]| > \varepsilon\right] = 0$$

根据概率测度 P 的连续性，这表明 $P[\lim_{n\to\infty}\max_{k\geqslant 0}|X[m+k]|X[m]>\varepsilon]=0$（参见定理8.1-1的引理）。最后，根据柯西收敛准则，存在一个随机变量 X 满足

$$X[n] \to X \quad \text{(a.s.)}$$

定理8.8-5 （强大数定理）令 $X[n]$ 为定义在 $n\geqslant 1$ 上的 WSS 独立随机序列，均值为 μ_X，方差为 σ_X^2，则当 $n\to\infty$

$$\hat{\mu}_X[n] = \frac{1}{n}\sum_{k=1}^{n} X[k] \to \mu_X \quad \text{(a.s.)}$$

证明 令 $Y[n] \triangleq \sum_{k=1}^{n}\frac{1}{k}X_c[k]$；则当 $n\geqslant 1$ 时 $Y[n]$ 为鞅序列。因为

$$E\{Y^2[n]\} = \sum_{k=1}^{n}\frac{1}{k^2}\sigma_X^2 \leqslant \sigma_X^2\sum_{k=1}^{\infty}\frac{1}{k^2} = C < \infty$$

可以应用定理8.8-4证明对某个随机变量 Y，$Y[n]\to Y$。又注意到 $X_c[k]=k(Y[k]-Y[k-1])$，可以得到

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^{n}X_c[k] &= \frac{1}{n}\left[\sum_{k=1}^{n}kY[k] - \sum_{k=1}^{n}kY[k-1]\right] \\ &= -\frac{1}{n}\sum_{k=1}^{n}Y[k] + \frac{n+1}{n}Y[n] \\ &\to -Y+Y=0 \qquad \text{(a.s.)} \end{aligned}$$

因此

$$\hat{\mu}_X[n] \to \mu_X \quad \text{(a.s.)}$$

小结

本章介绍了随机序列的概念，并且研究了其性质和描述方法。将随机序列定义为与样本空间中的实验结果或点相关的样本序列集。介绍了几种重要的随机序列。然后回顾了线性离散时间理论，考虑了样本序列通过系统后如何变换的实际问题。主要强调均值和协方差函数通过线性系统的变换。然后又研究了一些特殊并且重要的平稳和 WSS 随机序列的例子，引入了功率谱密度的概念。研究了随机序列的收敛以及评估各种可能的收敛方式。然后应用这些结论到大数定理中，并且应用鞅序列的性质证明了重要的强大数定理。

关于本章内容的一些其他资料可参阅参考文献[8-9]、参考文献[8-10]以及参考文献[8-11]。

下一章随着继续展开对随机过程的研究，我们将会看到本章的许多结论可以扩展到连续时间的情况。

习题

(标注星号 * 的习题带有一定的难度，需要花更多的时间和/或更深入的学习。)

8.1 证明 N 个事件交集的概率法则。例如，当 $N=3$，有

$$P[A_1A_2A_3]=P[A_1]P[A_2|A_1]P[A_3|A_1A_2]$$

并说明此法则对联合累积分布函数和联合概率密度函数有效。

*8.2 考虑一个 N 维随机向量 $\boldsymbol{X}$。验证此随机变量的元素两两独立并不表示这些元素是联合独立的。

8.3 令 $\boldsymbol{X}=[X_1, X_2, \cdots, X_5]^{\mathrm{T}}$为随机向量，其元素满足等式

$$X_i = X_{i-1} + B_i, \quad 1 \leqslant i \leqslant 5$$

其中 B_i 满足伯努利分布且联合独立，取值为 0 和 1，均值为 1/2。第一个值为 $X_1=B_1$。将 B_i 聚合在一起产生随机向量 $\boldsymbol{B}$。

(a) $\boldsymbol{A}$ 为某常数矩阵，记 $\boldsymbol{X}=\boldsymbol{AB}$，确定 $\boldsymbol{A}$。

(b) 求出均值向量 $\boldsymbol{\mu_X}$。

(c) 求出协方差矩阵 $\boldsymbol{K_{BB}}$。

(d) 求出协方差矩阵 $\boldsymbol{K_{XX}}$。

[对于(b)到(d)，用矩阵 $\boldsymbol{A}$ 来描述你的答案]。

8.4 令一簇序列 $x(n, \theta_k)$由一确定参数 θ_k 给出

$$\left\{\cos\left(\frac{2\pi n}{5} + \theta_k\right)\right\}_{k=0}^{N-1}$$

现定义在相同的参数集合$\{\theta_k\}$上取值的随机变量 Θ。令 Θ 的概率分布函数为

$$P_{\Theta}(\theta_k) = \frac{1}{N}, \qquad k=0,\cdots,N-1$$

现设 $X[n] \triangleq \cos\left(\frac{2\pi n}{5}+\Theta\right)$。

(a) $X[n]$是否为随机序列？如果是，描述映射 $X(n, \zeta)$及其概率空间$(\boldsymbol{\Omega}, \boldsymbol{F}, \boldsymbol{P})$。如果不是，给出完整解释。

(b) 令 $\theta_k=\frac{2\pi k}{N}$, $k=0, \cdots, N-1$, 求 $E\{X[n]\}$①。

(c) 对于与(b)中相同的 θ_k, 求 $E\{X[n]X[m]\}$。此处取 $N>2$。

*8.5 通常一个题目的开始如下:"令 X 为一实值随机变量, 其概率密度函数 $f_X(x)\cdots$"因为一个随机变量是源自具有事件域 $\mathscr{F}$ 和概率测度 P 的样本空间 Ω 的映射, 显然在问题的描述中假设了概率空间 $(\Omega, \mathscr{F}, P)$ 的存在性。论证总是可以创建适当的概率空间 $(\Omega, \mathscr{F}, P)$, 从而使如上的问题阐述合理。

8.6 令 T 为一连续随机变量, 表示第一个光子从光源中发出的时间; T 从光源加电的时刻开始计量。假设 T 的概率密度函数为 $f_T(t)=\lambda e^{-\lambda t}u(t)$, 其中 $\lambda>0$。

(a) 如果知道没有光子在 t_1 时刻之前发出, 则至少一个光子在 t_2 时刻之前发出的概率是多少, 其中 $t_1<t_2$?

(b) 如果有三个独立的光源同时加电, 则至少一个光子在 t_2 时刻之前发出的概率是多少?

8.7 设 X 为一个条件正态随机变量, 条件密度函数为 $N(\mu,\sigma^2)$, 给定随机变量的值 $M=\mu$ 以及 $\Sigma^2=\sigma^2$。

(a)假设 σ^2 是一个已知常量, M 是一个随机变量, 其累积分布函数为:

$$F_M(m) = [1-e^{-\lambda m}]u(m)\text{(注意到此处的变量 } m \text{ 是连续的!)}$$

其中 λ 是一个已知的正数。确定 X 的特征函数(提示:首先确定条件特征函数)。

(b)现在假设 Σ^2 和 M 都是独立的随机变量。设它们的分布任意, 但假设它们都有一个有限均值和方差。确定 X 的均值和方差, 用 μ_M, μ_{Σ^2}, σ_M^2 表示。

8.8 令 X 和 Y 为独立同分布随机变量, 其指数概率密度函数为

$$f_X(\omega)=f_Y(\omega)=\lambda e^{-\lambda w}u(\omega)$$

(a) 为以下分式确定概率密度函数 $f_R(r)$, $0<r\leqslant 1$

$$0\leqslant R\triangleq\frac{X}{X+Y}\leqslant 1$$

(b) 令 A 为事件 $X<1/Y$。假定 A 发生且 $Y=y$, 请确定 X 的条件概率密度函数; 也就是确定

$$f_X(x|A,Y=y)$$

(c) 利用(b)中的定义, 假定 A 发生且 $Y=y$, X 的最小均方误差估计是多少?

8.9 利用复值随机变量的施瓦茨不等式证明, 对任意复值 WSS 随机序列 $X[m]$ 有

$$|R_X[m]|\leqslant R_X[0],\text{对所有整数 } m$$

8.10 令 $\boldsymbol{X}=[X_1, X_2, \cdots, X_{10}]^{\mathrm{T}}$ 为一随机向量, 其元素满足方程

$$X_i=\frac{2}{5}(X_{i-1}+X_{i+1})+W_i,\qquad 2\leqslant i\leqslant 9$$

其中 W_i 是独立拉普拉斯分布, 均值为零, 方差为 σ^2, $i=1, 2, \cdots, 10$, 且 $X_1=\frac{1}{2}X_2+\frac{5}{4}W_1$ 及 $X_{10}=\frac{1}{2}X_9+\frac{5}{4}W_{10}$。

(a) 求均值向量 $\boldsymbol{\mu_X}$。

(b) 求协方差矩阵 $\boldsymbol{K_{XX}}$。

(c) 写出随机向量 $\boldsymbol{W}$ 的多维概率密度函数的表达式。

提示: 矩阵等式

若

① 注意: $\cos(A+B)=\cos A\cos B-\sin A\sin B$ 及 $\cos A\cos B=\frac{1}{2}\{\cos(A+B)+\cos(A-B)\}$。

$$A \triangleq \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^N \\ \rho & 1 & \rho & \rho^2 & \dots \\ \rho^2 & \rho & 1 & \dots & \dots \\ \dots & \rho^2 & \dots & \dots & \rho \\ \rho^N & \dots & \dots & \rho & 1 \end{bmatrix}$$

则 A^{-1} 为

$$\beta^2 A^{-1} = \begin{bmatrix} 1-\rho\alpha & -\alpha & 0 & \dots & 0 \\ -\alpha & 1 & -\alpha & 0 & \dots \\ 0 & -\alpha & 1 & \dots & 0 \\ \dots & 0 & \dots & \dots & -\alpha \\ 0 & \dots & 0 & -\alpha & 1-\rho\alpha \end{bmatrix}$$

其中 $\alpha \triangleq \dfrac{\rho}{1+\rho^2}$ 及 $\beta^2 \triangleq \dfrac{1-\rho^2}{1+\rho^2}$。拉普拉斯概率密度函数为

$$f_W(\omega) = \frac{1}{\sqrt{2}\sigma}\exp\left(-\sqrt{2}\frac{|\omega|}{\sigma}\right), \qquad -\infty \leqslant \omega \leqslant +\infty$$

8.11 证明推论 8.1-1。

8.12 令 $\{X_i\}$ 为独立同分布正态随机变量，均值为零，方差为 1。令

$$S_k \triangleq X_1 + X_2 + \cdots + X_k, \qquad k \geqslant 1$$

确定 S_n 和 S_m 的联合概率密度函数，其中 $1 \leqslant m < n$。

8.13 在例 8.1-8 中看到累积分布函数是右连续的。它们是否左连续？证明或给出反例。

8.14 令概率空间 $(\Omega, \mathscr{F}, P)$ 如下给定：$\Omega = \{a, b, c\}$，即输出 $\zeta = a$ 或 b 或 c；$\mathscr{F}$ 为所有 Ω 的子集；对每种输出 ζ，$P[\{\zeta\}] = \dfrac{1}{3}$。令随机序列 $X[n]$ 定义如下：

$$X[n,a] = 3\delta[n]$$

$$X[n,b] = u[n-1]$$

$$X[n,c] = \cos \pi n/2$$

(a) 求均值函数 $\mu_X[n]$。

(b) 求相关函数 $R_{XX}[m, n]$。

(c) $X[1]$ 和 $X[0]$ 是否独立？为什么？

8.15 设平稳高斯随机序列 $Y[n]$ 具有零均值，协方差函数为

$$K_{YY}[m] = \sigma^2 \rho^{|m|} \quad 其中 \quad -\infty < m < +\infty, |\rho| < 1$$

(a) 给定 $X[n-1]$，求 $X[n]$ 的条件均值。

(b) 什么意义下它是 $X[n]$ 的一个好的预测。

8.16 考虑一随机序列 $X[n]$ 输入具有以下冲激响应的线性滤波器

$$h[n] = \begin{cases} 1/2, & n = 0 \\ 1/2, & n = 1 \\ 0, & 其他 \end{cases}$$

记输出随机序列 $Y[n]$，即对每个输出 ζ

$$Y[n,\zeta] = \sum_{k=-\infty}^{k=+\infty} h[k] X[n-k,\zeta]$$

假设此滤波器一直运行，即 $-\infty < n < +\infty$。已知输入均值函数 $\mu_x[n]$ 和相关函数 $R_{XX}[n_1, n_2]$，用这些假设已知的函数回答以下问题：

(a) 求输出的均值函数 $\mu_Y[n]$。

(b) 求输出的自相关函数 $R_{YY}[n_1, n_2]$。

(c) 用(a)和(b)的解答表示输出的自协方差函数 $K_{YY}[n_1, n_2]$。

(d) 假设输入 $X[n]$ 是高斯随机序列，用 $\mu_Y[n]$ 和 $K_{YY}[n_1, n_2]$ 表示输出在任意两个时刻 $n_1 \neq n_2$ 的联合概率密度函数 $f_Y(y_1, y_2; n_1, n_2)$。

8.17 随机到达时间序列 $T[n]$，定义在 $n \geqslant 1$，具有 Erlang 概率密度，对某些 $\lambda > 0$

$$f_T(t;n) = \frac{(\lambda t)^{n-1}}{(n-1)!}\lambda \exp(-\lambda t)u(t)$$

求联合概率密度函数 $f_T(t_2, t_1; 10, 5)$。其中 $T[n]$ 是指数概率密度函数的独立同分布到达时间序列直到 n 时刻之前的所有序列之和。

8.18 令 $T[n]$ 表示随机到达序列

$$T[n] = \sum_{k=1}^{n} \tau[k]$$

其中 $\tau[k]$ 为到达间隔时间的独立随机序列，满足 $\lambda > 0$ 的指数分布。

(a) 求此随机序列的特征函数，即

$$\Phi_T(\omega; n) = E[\mathrm{e}^{\mathrm{j}\omega T[n]}]$$

(b) 利用此特征函数求均值函数 $\mu_T[n]$。

8.19 令随机序列 $T[n]$ 定义在 $n \geqslant 1$ 上，且对每个 n 具有 Erlang 概率密度函数

$$f_T(t;n) = \frac{(\lambda t)^{n-1}}{(n-1)!}\lambda \mathrm{e}^{-\lambda t}u(t), \quad \lambda > 0$$

对 $n \geqslant 2$ 定义新的随机序列 $\tau[n] \triangleq T[n] - T[n-1]$，且设 $\tau[1] \triangleq T[1]$。可否得出结论 $\tau[n]$ 为具有相同参数 λ 的指数分布？如果不是，需要什么关于 $T[n]$ 的额外信息？验证你的答案。

8.20 考虑一个具有“漏电”电池的充电耦合器(CCD)阵列的随机序列。首先定义宽度 3 的脉冲函数

$$h[n] = \begin{cases} 1/4, & n = -1 \\ 1/2, & n = 0 \\ 1/4, & n = +1 \\ 0, & \text{其他} \end{cases}$$

如图 P8.20 所示，表示25% 的电量泄漏到右边，25% 的电量泄漏到左边。假设一维的 CCD 阵列是无限长的，用随机序列 X 表示阵列的无限维试验输出

$$X[n,\zeta] = \sum_{i=-\infty}^{i=+\infty} A(\zeta_i)h[n-i]$$

其中 ζ_i 是 ζ 的第 i 个组件。随机变量 $A(\zeta_i)$ 为联合独立的高斯分布，均值和方差均为 λ。

(a) 求均值函数 $\mu_X[n]$。

(b) 求一阶概率密度函数 $f_X(x; n)$。

(c) 求联合概率密度函数 $f_X(x_1, x_2, n, n+1)$。

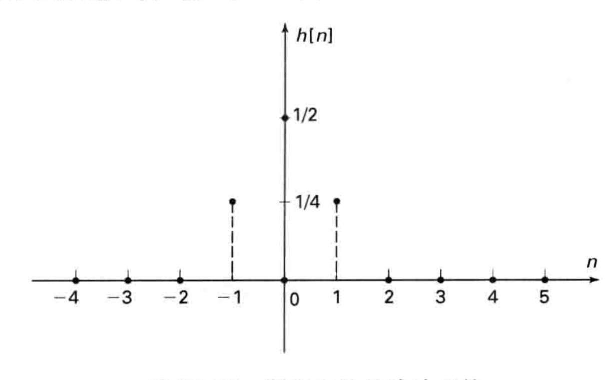


图 P8.20 漏电电池的脉冲函数

8.21　给定随机序列 $X[n]$，$n \geqslant 0$，其条件概率密度函数为

$$f_X(x_n|x_{n-1}) = \alpha \exp[-\alpha(x_n - x_{n-1})]\ u(x_n - x_{n-1}), \qquad n \geqslant 1$$

其中 $u(x)$ 是单位阶梯函数，初始概率密度函数为 $f_X(x_0)=\delta(x_0)$，取 $\alpha>0$。

(a) 求 $n=2$ 时一阶概率密度函数 $f_X(x_n)$。

(b) 利用数学归纳法求任意 $n>1$ 的一阶概率密度函数 $f_X(x_n)$。

8.22　令 $X[n]$ 为如图 P8.22 所示的离散时间 LSI 系统 H 的确定性输入。

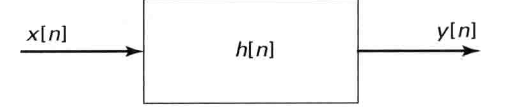
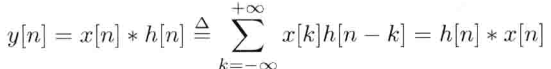


图 P8.22　具有冲激响应 $h[n]$ 的 LSI 系统

(a) 利用线性性和移不变性证明



$$y[n] = x[n] * h[n] \triangleq \sum_{k=-\infty}^{+\infty} x[k]h[n-k] = h[n] * x[n]$$

(b) 定义序列 $a[n]$ 的傅里叶变换为

$$A(\omega) \triangleq \sum_{n=-\infty}^{\infty} a[n]\mathrm{e}^{-\mathrm{j}\omega n}, \qquad -\pi \leqslant \omega \leqslant +\pi$$

试证明其傅里叶逆变换为

$$a[n] = \frac{1}{2\pi}\int_{-\pi}^{+\pi} A(\omega)\mathrm{e}^{+\mathrm{j}\omega n}\mathrm{d}\omega, \qquad -\infty < n < +\infty$$

(c) 利用(a)和(b)的结果证明对于 LSI 离散时间系统，有

$$Y(\omega) = H(\omega)X(\omega), \qquad -\pi \leqslant \omega \leqslant +\pi$$

8.23　考虑差分方程

$$y[n] + \alpha y[n-1] = x[n], \qquad -\infty < n < +\infty$$

其中 $-1<\alpha<+1$。

(a) 令输入为 $x[n]=\beta^n u[n]$ $(-1<\beta<+1)$。假设因果性成立，即对 $n<0$，$y[n]=0$，求 $y[n]$ 的解。

(b) 令输入为 $x[n]=\beta^{-n}u[-n]$ $(-1<\beta<+1)$。假设反因果性成立①，即对 $n>0$，$y[n]=0$，求 $y[n]$ 的解。

8.24　令 $X[n]$ 为 WSS 随机序列，均值为零，且协方差函数

$$K_{XX}[m] = \sigma^2\rho^{|m|}, \quad \text{对于所有} -\infty < m < +\infty$$

其中 ρ 是实常数。考虑以下形式的差分方程

$$Y[n] = X[n] - \alpha X[n-1], \qquad \text{且} -\infty < n < +\infty$$

(a) 用参数 σ^2，ρ，α 表示 $Y[n]$ 的协方差函数。

(b) 求 α 使 $Y[n]$ 为 WSS 白噪声序列。

(c) 求此白噪声的平均功率。

8.25　令 $W[n]$ 为独立随机序列，均值为 0，方差为 σ_W^2，对 $-\infty<n<+\infty$ 有定义。对适当选择的 ρ，使稳态随机序列 $X[n]$ 满足因果 LCCDE

$$X[n] = \rho X[n-1] + W[n], \qquad -\infty < n < +\infty$$

(a) 证明 $X[n-1]$ 和 $W[n]$ 在时刻 n 是独立的。

(b) 推导出特征函数方程

$$\Phi_X(\omega) = \Phi_X(\rho\omega)\Phi_W(\omega)$$

(c) 对于未知函数 Φ_X，当 $W[n]$ 假设为高斯的，求此函数方程的连续解(注意：$\Phi_X(0)=1$)。

(d) σ_X^2 是多少?

① 这部分要求掌握关于 z 变换的详细知识(参见附录 A)。

8.26　考虑图 P8.26 所示的 LSI 系统，其确定输入 $x[n]$ 被随机噪声序列 $W[n]$ 污染。希望确定输出随机序列 $Y[n]$ 的性质。噪声 $W[n]$ 的均值 $\mu_W[n]=2$，其自相关函数为 $E\{W[m]W[n]\}=\sigma_W^2\delta[m-n]+4$。冲激响应为 $h[h]=\rho^n u[n]$，$|\rho|<1$。确定输入 $x[n]$ 对所有 n 均为 3。

(a) 求输出的均值 $\mu_Y[n]$。

(b) 求输出功率 $E\{Y^2[n]\}$。

(c) 求输出的协方差 $K_{YY}[m,n]$。

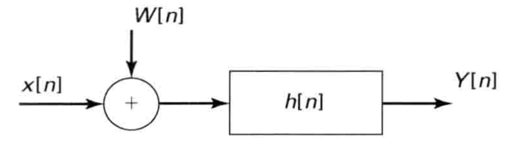


图 P8.26　确定信号加噪声输入的 LSI 系统

8.27　证明例 8.1-15 产生的随机序列 $X[n]$ 不是独立随机序列。

8.28　设 $W[n]$ 为一个独立随机序列，均值 $\mu_W=0$，方差为 σ_W^2。定义一个新的随机序列如下：

$$X[0]=0$$

$$X[n]=\rho X[n-1]+W[n],\quad n\geqslant 1$$

(a) 求 $X[n]$ 的均值，$n\geqslant 0$。

(b) 求 $X[n]$ 的协方差，$K_{XX}[m,n]$。

(c) ρ 为何值，则当 m 和 n 变大时，$K_{XX}[m,n]$ 趋近于 $G[m-n]$（对某些有限值函数 G）？这种情况被称为渐进平稳。

8.29　设随机变量 A 和 B 为独立同分布，均值为 0，方差 σ^2，三阶矩 $E[A^3]=E[B^3]=m_3\neq 0$。考虑随机序列 $X[n]=A\cos\omega_0 n+B\sin\omega_0 n,\ -\infty<n<+\infty$，其中 ω_0 是一个固定的射频频率。

(a) 证明 $X[n]$ 是 WSS。

(b) 通过举出一个反例，证明 $X[n]$ 非平稳。

8.30　考虑一 WSS 随机序列 $X[n]$，均值 $\mu_X[n]=\mu$ 为常数，其相关函数为 $R_{XX}[m]=p^2\delta[m]$ 且 $p^2>0$。在这种情况下，μ 必为 0（此题中将得到证明）。注意本例中协方差函数为 $K_{XX}[m]=p^2\delta[m]-\mu^2$。

(a) 取 $m=0$，证明 $p^2\geqslant\mu^2$。

(b) 从随机序列 $X[n]$ 中取长度为 N 的向量 $\boldsymbol{X}$。证明相应的协方差矩阵 $\boldsymbol{K_{XX}}$ 只有当 $\mu^2\leqslant\sigma^2/(N-1)$ 时是半正定的，其中 $\sigma^2\triangleq p^2-\mu^2$（提示：取系数向量 $\boldsymbol{a}=\boldsymbol{1}$，即都为 1）。

(c) 令 $N\to\infty$ 并证明对于平稳白噪声序列 $X[n]$，μ 必为 0。

8.31　对习题 8.16 中的线性滤波器，假设输入随机序列是 WSS，将输出功率谱密度 $S_{YY}(\omega)$ 用输入功率谱密度 $S_{XX}(\omega)$ 来表示。

(a) 证明功率谱密度是一个实函数，即使 $X[n]$ 是一个复值随机序列。

(b) 证明如果 $X[n]$ 是实值，则 $S_{XX}(\omega)=S_{XX}(-\omega)$。

(c) 证明 $S_{XX}(\omega)\geqslant 0$ 对所有 ω 成立，无论 $X[n]$ 是否为复值。

8.32　令 WSS 随机序列 X 的相关函数

$$R_{XX}[m]=10\mathrm{e}^{-\lambda_1|m|}+5\mathrm{e}^{-\lambda_2|m|}$$

其中 $\lambda_1>0$ 及 $\lambda_2>0$。求 $|\omega|\leqslant\pi$ 时，相应的功率谱密度 $S_{XX}(\omega)$。

8.33　某随机序列的功率谱密度为 $S_{XX}(\omega)=1/[(1+\alpha^2)-2\alpha\cos\omega]^2$，$-\pi\leqslant\omega\leqslant+\pi$，其中 $|\alpha|<1$。求随机序列的相关函数 $R_X[m]$。

8.34　令系统 $H(\omega)$ 的输入 $W[n]$ 为一白噪声随机序列，$\mu_W[n]=0$ 且 $K_{WW}[m]=\delta[m]$。令 $X[n]$ 表示相应的输出随机序列。求 $K_{XW}[m]$ 及 $S_{XW}(\omega)$。

8.35　考虑图 P8.35 所示的系统。令 $X[n]$ 和 $V[n]$ 为 WSS 且不相关，均值都为 0，功率谱密度分别为 $S_{XX}(\omega)$ 和 $S_{VV}(\omega)$。

(a) 求输出的功率谱密度 $S_{YY}(\omega)$。

(b) 求输入 X 和输出 Y 的互功率谱密度，即 $S_{XY}(\omega)$。

8.36　考虑离散时间系统，其输入为随机序列 $X[n]$，输出 $Y[n]$ 为

$$Y[n] = \frac{1}{5}\sum_{k=-2}^{+2} X[n-k]$$

假设输入序列 $X[n]$ 为 WSS，且其功率谱密度 $S_{XX}(\omega)=2$。

(a) 求输出随机序列的功率谱密度 $S_{YY}(\omega)$。

(b) 求输出相关函数 $R_{YY}[m]$。

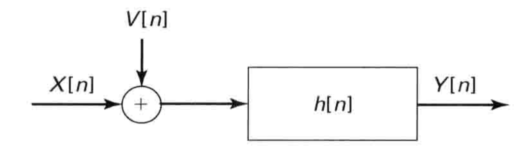


图 P8.35　具有随机信号加噪声输入的 LSI 系统

8.37　令平稳随机序列 $Y[n]=X[n]+U[n]$ 为特定离散时间信道信号 X 和噪声 U 的模型，其功率谱密度 $S_Y(\omega)$。假定 X 和 U 是正交的，又假定对所有 $|\omega|\leqslant\pi$ 都有 $S_Y(\omega)>0$。作为通过 Y 获得对 X 的估计的第一步，令 Y 输入一离散时间滤波器 $G(\omega)$，定义为 $G(\omega)=1/\sqrt{S_Y(\omega)}$，产生如图 P8.37(a)所示的平稳输出序列 $W[n]$。

(a) 基于 S_X, S_Y, S_U 求 $W[n]$的功率谱密度，即 $S_W(\omega)$，及原始输入与输出之间的互功率谱密度$S_{XW}(\omega)$。

(b) 接下来用一个冲激响应 $h[n]$，$n=0, 1, \cdots, N-1$ 的 FIR 对 $W[n]$进行滤波，输出 $\hat{X}[n]$，原无噪声信号 $X[n]$的估计如图 P8.37(b)所示。按照随机变量的希尔伯特空间理论确定选择滤波器的系数 $h[n]$，则估计误差 $\hat{X}[n]-X[n]$将与所有时刻 n 用来估计的 $W[n]$相互正交。写出方程的 N 个滤波器系数 $h[0]$, $h[1]$, $\cdots$, $h[N-1]$。答案需要用互相关函数 $R_{XW}[m]$表示。

(c) 令 N 趋向无穷，写出 $h[n]$的频率响应，即 $H(\omega)$，用离散时间功率谱密度 $S_{XX}(\omega)$和 $S_{YY}(\omega)$表示。

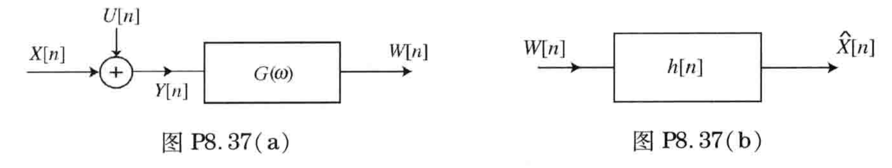


图 P8.37(a)　　　　图 P8.37(b)

8.38　二阶矩及更高阶矩已经证明在许多高级应用中非常有用。下面考虑平稳随机序列 $x[n]$，$-\infty<n<+\infty$ 的三阶相关函数：

$$R_X[m_1,m_2] \triangleq E\{X[n+m_1]X[n+m_2]X^*[n]\}$$

(a) 令 $Y[n]$为冲激响应 $h[n]$的 LSI 系统的输出，作为输入随机序列 $X[n]$的响应。确定一个类似卷积的方程表示输出的三阶相关函数 $R_Y[m_1, m_2]$，用输入的三阶相关函数 $R_X[m_1, m_2]$和系统的冲激响应 $h[n]$表示。定义 X 的双频谱密度为二维傅里叶变换

$$S_X(\omega_1,\omega_2) = \sum_{m_1}\sum_{m_2} R_X[m_1,m_2]\exp-\mathrm{j}(\omega_1 m_1+\omega_2 m_2)$$

(b) 对于(a)中的系统，求输出的双频谱密度 $S_Y(\omega_1, \omega_2)$，用系统的频率响应 $H(\cdot)$和输入的双频谱密度 $S_X(\omega_1, \omega_2)$表示。

8.39　令 $X[n]$为 $n\geqslant0$ 上的马尔可夫链，取值为 1 和 2，其一步转移概率为

$$\boldsymbol{P}_{ij} \triangleq P\{X[n]=\mathrm{j}|X[n-1]=i\}, \qquad 1\leqslant i,j\leqslant 2$$

其矩阵形式为

$$\boldsymbol{P} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = (p_{i,j})$$

用向量表示 n 时刻的状态概率

$$\boldsymbol{p}[n] \triangleq [P\{X[n]=1\}, P\{X[n]=2\}]$$

(a) 证明$\boldsymbol{p}[n]=\boldsymbol{p}[0]\boldsymbol{P}^n$。

(b) 绘出二状态转移图，并在每个分支上标注一步转移概率p_{ij}。不要忘记p_{ii}或自转移(参见图8.5-1的马尔可夫链状态转移图)。

(c) 给定$X[0]=1$，求时刻n发生的首先转移到状态2的概率。

8.40 考虑采用一阶马尔可夫序列建模随机序列$X[n]$如下：

$$X[n]=rX[n-1]+Z[n]$$

其中$Z[n]$是方差为σ_Z^2的白噪声。取$|r|<1$且假设系统已经长时间运行，即$-\infty<n<+\infty$。

(a) 求$X[n]$的功率谱密度，即$S_{XX}(\omega)$。

(b) 求相关函数$R_{XX}[m]$。

8.41 本章定义的马尔可夫随机序列$X[n]$是用其一阶概率密度函数$f_X(x;\ n)$及其一步条件概率密度函数

$$f_X(x_n|x_{n-1};n,n-1)=f_X(x_n|x_{n-1})$$

(a) 用上面的函数表示马尔可夫随机序列的两步概率密度函数$f_X(x_n|x_{n-2})$。这里对于起始于$n=0$的随机序列，取$n\geqslant 2$。

(b) 对任意正整数N，求N步概率密度函数$f_X(x_n|x_{n-N})$，只需要考虑$n\geqslant N$。

8.42 考虑在$\{n\geqslant 0\}$上一个随机游动序列$X[n]$，定义如下：

$$X[0]\triangleq 0$$

$$X[n]\triangleq\sum_{k=1}^{n}W[k],\quad n\geqslant 0$$

其中$W[n]$为独立随机序列，平稳，且满足

$$W[n]\triangleq\begin{cases}+s_1, p=1/2\\ -s_2, p=1/2\end{cases}$$

差别仅在于正负间距大小$s_1\neq s_2$，且$s_1>0$，$s_2>0$

(a)求均值函数$\mu_X[n]\triangleq E\{X[n]\}$。

(b)求自相关函数$R_X[n_1,n_2]\triangleq E\{X[n_1]X[n_2]\}$。

8.43 考虑一个马尔可夫随机序列$X[n]$，其中$1\leqslant n\leqslant 100$。其一阶概率密度为$f_X(x;\ 1)$，一步转移概率密度为$f_X(x_n|x_{n-1};\ n,\ n-1)$。由马尔可夫的定义，我们有

$$f_X(x_n|x_{n-1})=f_X(x_n|x_{n-1},x_{n-2},\cdots,x_1),\qquad 2\leqslant n\leqslant 100$$

说明一个马尔可夫随机序列倒序后仍为马尔可夫序列，即

$$f_X(x_n|x_{n+1})=f_X(x_n|x_{n+1},x_{n+2},\cdots,x_{100}),\quad 1\leqslant n\leqslant 99$$

这样描述这个马尔可夫随机序列的统计特性也可用一步反向概率密度$f_X(x_{n-1}|x_n;\ n-1,\ n)$和一阶概率密度$f_X(x;\ 100)$来描述。

8.44 给定一个马尔可夫链$X[n]$，$n\geqslant 1$，转移概率为$P[x[n]|x[n-1]]$，求两步转移概率$P[x[n]|x[n-2]]$的表达式。并证明

$$P[x[n+1]|x[n-1],x[n-2],\cdots,x[1]]=P[x[n+1]|x[n-1]],\quad n\geqslant 1$$

8.45 考虑一个$n\geqslant 1$的马尔可夫随机序列$X[n]$，由差分方程产生

$$X[n]=\alpha X[n-1]+\beta W[n]$$

其中输入$W[n]$为独立随机序列，均值为0，方差为σ_W^2，初始值$X[0]=0$，参数α和β为已知常数。

(a)说明子序列$Y[n]\triangleq X[2n]$也为马尔可夫序列。

(b)求$n>0$时的方差函数$\sigma_Y^2[n]\triangleq E[|Y[n]-\mu_Y[n]|^2]$。

*8.46 写一个MATLAB函数`triplemarkov`用来计算和绘制例8.1-16中不对称的两状态马尔可夫模型在任意三组参数$\{p_{00},p_{11}\}$时的自相关函数。设定最大间隔为N，在$\{0.2, 0.8\}$，$\{0.2, 0.5\}$和$\{0.2, 0.2\}$

时运行程序。在{0.8, 0.2}, {0.8, 0.5}和{0.8, 0.8}时再运行一次。分析观察到的现象。

8.47 考虑概率空间(Ω, $\mathscr{F}$, P), 其中 $\Omega=[0,1]$, $\mathscr{F}$定义为Ω的波雷尔集, 且 $P[0,\zeta]=\zeta$, $0<\zeta\leqslant 1$。

(a)用概率原理说明 $P[\{0\}]=0$。

(b)确定在什么意义下以下随机序列收敛:

(i) $X[n,\zeta]=\mathrm{e}^{-n\zeta}, n\geqslant 0$

(ii) $X[n,\zeta]=\sin\left(\zeta+\frac{1}{n}\right), n\geqslant 1$

(iii) $X[n,\zeta]=\cos^n(\zeta), n\geqslant 0$

(c)如果上述序列收敛, 则其板限值为多少?

8.48 相互独立的随机变量 $X[n]$的各元素的概率密度具有如下形式:

$$f_X(x;n)=\left(1-\frac{1}{n}\right)\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}\left(x-\frac{n-1}{n}\sigma\right)^2\right]+\frac{1}{n}\sigma\exp(-\sigma x)u(x)$$

确定下述情况下随机序列 $X[n]$是否收敛。

(i)在均方意义下

(ii)概率

(iii)分布

8.49 随机序列 $X[n]$中元素的联合概率密度具有如下形式:

$$f_X(\alpha,\beta;m,n)=\frac{mn}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}[m^2\alpha^2-2\rho mn\alpha\beta+n^2\beta^2]\right)$$

满足 $m\geqslant 1$ 和 $n\geqslant 1$, 其中 $-1<\rho<+1$。

(a)证明 $X[n]$当 $n\to\infty$ 时对所有 $-1<\rho<+1$ 在均方意义下收敛。

(b)确定均方极限 $X\triangleq\lim_{n\to\infty}X[n]$的 CDF。

8.50 列出一个高斯随机变量序列的均方极限仍为高斯条件。

8.51 设 $X[n]$为 $n\geqslant 0$ 的实随机序列, 组成平稳的、独立的增量, 即 $X[n]-X[n-1]=W[n]$, "增量" $W[n]$为一个平稳独立的随机序列。随机序列总是起始于 $X[0]=0$。我们还知道在 $n=1$ 时, $E\{X[1]\}=\eta$, $\mathrm{Var}\{X[1]\}=\sigma^2$。

(a)求 $\mu_X[n]$ 和 $\sigma_X^2[n]$, 随机序列 X 在 $n(n>1)$时刻的均值和方差函数。

(b)证明 $X[n]/n$ 当 n 趋向于无穷大时依概率收敛于 η。

8.52 这个问题论证 p 收敛说明了在分布意义上收敛, 即使当极限概率密度不存在。

(a)对任意实数 x 和任意正数 ε, 证明

$$P[X\leqslant x-\varepsilon]\leqslant P[X[n]\leqslant x]+P[|X[n]-X|\geqslant\varepsilon]$$

(b)同样证明

$$P[X>x+\varepsilon]\leqslant P[X[n]>x]+P[|X[n]-X|\geqslant\varepsilon]$$

(c)中假设随机序列 $X[n]$依概率收敛于随机变量 X。

(c)设 $n\to\infty$, 则在 F_X的连续点上有

$$\lim_{n\to\infty}F_X(x;n)=F_X(x)$$

8.53 设 $X[n]$为一个二阶随机序列, $h[n]$为 LSI 系统的冲激响应。我们希望定义系统输出 $Y[n]$为均方极限。

(a)证明我们可以定义均方极限为

$$Y[n]\triangleq\sum_{k=-\infty}^{+\infty}h[k]X[n-k],\qquad -\infty<n<+\infty,\ (\text{m.s.})$$

如果对所有 n 满足

$$\sum_k\sum_l h[k]h^*[l]R_{XX}[n-k,n-l]<\infty$$

(提示: 设 $Y_N[n]\triangleq\sum_{k=-N}^{+N}h[k]X[n-k]$, 用柯西收敛准则来证明 $Y_N[n]$的 m.s 极限存在)。

(b)当$X[n]$为一广义平稳随机序列时，找一个更简单的条件。

(c)当$X[n]$为(平稳)白噪声时，求关于$h[n]$的必要条件。

8.54 如果$X[n]$为一个定义在$n \geqslant 0$上的鞅序列，证明对所有$n \geqslant 0$有

$$E\{X[n+m]|X[m],\cdots,X[0]\} = X[m]$$

8.55 设$Y[n]$为一随机序列，X为一随机变量，考虑条件期望

$$E\{X|Y[0],\cdots,Y[n]\} \triangleq G[n]$$

证明随机序列$G[n]$为一个鞅序列。

*8.56 我们可以将鞅序列的定义扩展如下：设对每个$n \geqslant 0$，可测函数g有$G[n] \triangleq g(X[0], \cdots, X[N])$。如果

$$E\{G[n]X[0],\cdots,X[n-1]\} = G[n-1],则 G 为一个关于 X 的鞅序列$$

(a)证明根据定理8.8-3，G仍为关于X的鞅序列。特别地，在定理的表述中需要做些变化，即用G取代X。

(b)证明根据鞅收敛定理8.8-4，G仍为关于X的鞅序列。

8.57 考虑假设检验问题，对随机序列X的$(n+1)$次观测$X[0]$，…，$X[n]$，定义似然比

$$L_X[n] \triangleq \frac{f_X(X[0],\cdots,X[n]|H_1)}{f_X(X[0],\cdots,X[n]|H_0)}, \qquad n \geqslant 0$$

对应两种假设H_1和H_0。证明在H_0假设下，$L_X[n]$为一个关于X的鞅序列。

8.58 在例8.4-7中有关内插的讨论，解释在得到上采样的随机序列$X_e[n]$的功率谱密度所需要的代数关系式。

8.59 在内插过程中的上采样的随机序列$X_e[n]$很显然不是WSS，即使$X[n]$是WSS。通过随机化序列$X[n]$的起始时刻，可产生一个上采样的WSS随机序列。即定义一个二元随机变量Θ，满足$P[\Theta=0]=P[\Theta=1]=0.5$。用$X_r[n] \triangleq X[n+\Theta]$定义随机起始时刻的序列，则得到的上采样序列为$X_{er}[n]=X\left[\frac{n+\Theta}{2}\right]$。证明当$k$为奇数时$R_{X_rX_r}[k]=R_{XX}[k]$和$R_{X_{er}X_{er}}[m, m+k]=R_{X_{er}X_{er}}[k]=0.5R_{XX}[k/2]$，当$k$为偶数时为零。

参考文献

8-1 R. B. Ash, *Real Analysis and Probability.* New York: Academic Press, 1972, pp. 1-53 and 115.

8-2 A. Kolmogorov, *Foundations of the Theory of Probability.* New York: Chelsea, 1950.

8-3 T. M. Apostol, *Mathematical Analysis.* Reading, MA: Addison-Wesley, 1957, pp. 192-202.

8-4 Y. Viniotis, *Probability and Random Processes.* Boston, MA: WCB/Mcgraw-Hill, 1998, p. 103.

8-5 A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd Edition. Upper Saddle River, NJ: Prentice Hall, 1999, Chapters 2-3.

8-6 H. Stark and Y. Yang, *Vector Space Projections.* New York: Wiley, 1998.

8-7 S. S. Soliman and M. D. Srinath, *Continuous and Discrete Signals and Systems.* Upper Saddle River, NJ: Prentice Hall, 1998.

8-8 W. Rudin, *Principles of Mathematical Analysis.* New York: McGraw-Hill, 1964, pp. 45-48.

8-9 B. V. Gnedenko, *The Theory of Probability* (translated by B. D. Seckler). New York: Chelsea, 1963, p. 237.

8-10 A. Leon-Garcia, *Probability, Statistics, and Random processes for Electrical Engineering*, 3rd edition. Upper Saddle River, NJ: Prentice Hall, 2008.

8-11 G. Grimmett and D. Strizaker, *Probability and Random Processes*, 3rd edition. Oxford, England: Oxford University Press, 2001.

第9章 随机过程

上一章给出了从随机变量到随机序列的定义。只需将 $\zeta \in \Omega$ 的每个输出与一个样本序列对应起来，就得到了一簇样本序列，我们称其为随机序列。这些序列 Z 通过一个离散的(整数)参量 n 来排序。这一章我们考虑连续参量的随机函数。考察的连续参量可以是时间，也可以是位置或角度，或者其他的连续参量。所有这些连续时间函数的集合称为一个随机过程。随机过程可能是我们学习最有用的对象，因为它们可以用来直接对物理过程建模，而不需要对数据进行采样。甚至当一定要处理采样数据时，随机过程的概念也可以让我们对有限的连续过程套用采样序列的性质判断采样速率是否足够。

随机过程有着广泛的应用。其中应用最普遍的是作为物理系统中的噪声模型，要想减弱噪声带来的负面影响，最好的方法就是首先对其建模，第二类应用则对虽然不是噪声但对系统设计者来说仍然是未知的这类随机现象的建模。例如通信链路中的多媒体信号(音频、图像或视频)。这个信号虽然不是噪声，但对于一个具有一定距离的接收者来说它是未知的，并且可以取任意值。因此，当这些信号的统计类型是变化的时，我们用随机过程来对它们建模。这种情况在控制系统、模式识别等问题中也会产生。实际上，从信息论的角度看，通信系统中的波形也是具有一定随机性的。

我们从随机过程的定义开始学习，然后对处理连续时间时遇到的一些难点问题进行研究。然后对照这个随机过程的时间函数依据第8章学习的方法对这种连续参量的情况生成其相关函数和协方差函数。我们还将学习一些在实际中非常重要的基本随机过程。然后开始学习线性系统和随机过程。事实上，这是我们学习随机过程的核心问题，在实际中得到了广泛的应用。之后，根据随机过程其统计特性的不同对其进行分类。最后，介绍平稳和广义平稳随机过程及其线性系统分析。

9.1 基本定义

完整地理解随机过程的基本定义以及相关的其他函数是非常重要的。第8章分析的是将模拟信号转换成离散时间序列的情况，现在主要的新问题是时间轴为不可数的，我们将从基本定义开始讨论。

定义9.1-1 设 $(\Omega, \mathscr{F}, P)$ 为一概率空间。定义一个从样本空间 Ω 到一个连续时间函数空间的映射 X。这个空间中的元素被称为样本函数。如果对每一个固定的时刻 t，映射是一个随机变量，即对实轴 $-\infty < t < +\infty$ 上的每一个固定的 t，$X(t, \zeta) \in \mathscr{F}$ [①]，那么，这个映射就称为随机过程。

这样我们就有了一个多维函数 $X(t, \zeta)$，对应于每一个确定的实验结果 ζ，它是一个普通的时间函数，而对于某一确定时刻 t，它是一个随机变量。图9.1-1中样本空间 Ω 对应的是连续时间区间[0, 10]的特殊情况。如果沿时间轴观察，则得到一簇以时间 t 标注的随机变量，如果沿样本轴观察，则得到一簇以实验结果标注的时间函数。

---

① $X \in \mathscr{F}$ 是 $\{\zeta : X(\zeta) \leqslant x\} \subset \mathscr{F}$ 的缩写。这个条件允许我们对这类事件的概率进行测量，并因此引出了CDF的定义。

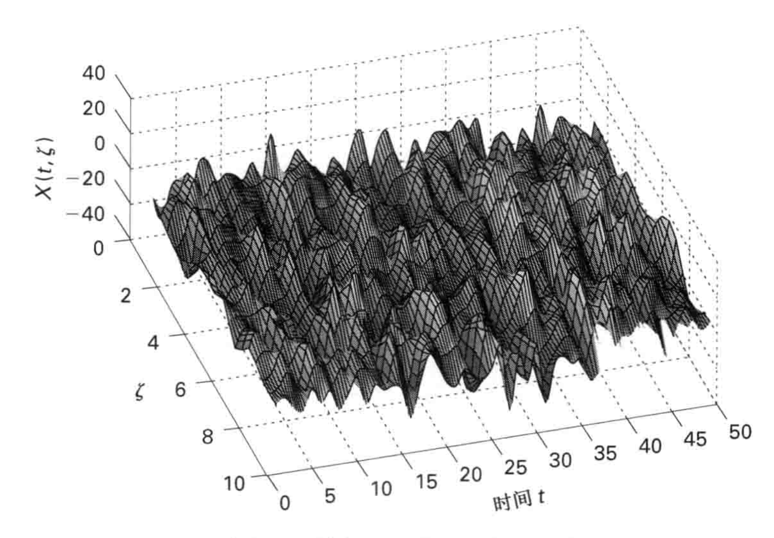


图9.1-1 对应连续样本 $\Omega=[0, 10]$ 的一个随机过程

以下列举一些随机过程的基本例子。

例 9.1-1 （简单过程）$X(t, \zeta)=X(\zeta)f(t)$，其中 X 是一个随机变量，f 是一个关于时间 t 的确定函数。一般将其表示为 $X(t)=Xf(t)$。

例 9.1-2 （随机正弦波）$X(t, \zeta)=A(\zeta)\sin(\omega_0 t+\Theta(\zeta))$，其中 A 和 Θ 是随机变量。一般将其表示为 $X(t)=A\sin(\omega_0 t+\Theta)$，省略了实验结果 ζ。

从随机序列中还可构造出很多随机过程的典型例子。

例 9.1-3 $X(t)=\sum_n X[n]p_n(t-T[n])$，其中 $X[n]$ 和 $T[n]$ 是随机序列，函数 $p_n(t)$ 是可具有任意形状的确定波形。例如，$p_n(t)$ 可以是理想的单位阶跃函数，这样就构造了一个被称为“跳过程”的模型。$T[n]$ 是到达时间，$X[n]$ 是跳跃的幅度。$X(t)$ 表示的是 t 之前所有时间的幅值。如果 $X[n]$ 始终为1，则 $X(t)$ 为一个计数过程，它将对先于 t 到达的数目进行计数。

如果对随机过程在 t_1 到 t_n 这 n 个时刻进行采样，可以得到一个 n 维随机向量。如果这个随机向量从 t_1 到 t_n 这 n 个时刻的概率分布已知，则显然对这个随机过程我们知之甚多。如果所有这些信息都知道，则说这个随机过程“统计特性”已知，与随机序列的情况类似。

定义 9.1-2 随机过程 $X(t)$ 是由它的 n 阶 CDF（pdf 或 PMF）$F_X(x_1, x_2, \cdots, x_n; t_1, t_2, \cdots, t_n)$ 的一个完备集来统计描述的，其中 n 为所有正整数，$x_1, x_2, \cdots, x_n$ 为实数，$-\infty<t_1<t_2<\cdots<t_n<+\infty$。

“统计”这个词来源于我们从积累相关的事件发生的频数中得到信息。显然，这是对我们希望建模的随机过程 $x(t)$ 通过测量预期能够得到的。但是，问题来了：这些信息就足以完全描述这个随机过程了吗？显然不是，我们需要一个连续的样本函数 $x(t)$，参见下面的例子。

例 9.1-4 （来源于 Karlin[9-1]）设 U 是服从$[0, 1]$之间均匀分布的随机变量，定义随机过程 $X(t)$ 和 $Y(t)$ 如下：

$$X(t) \triangleq \begin{cases} 1, & t=U \\ 0, & \text{其他} \end{cases}$$

$$Y(t) \triangleq 0 \quad \text{对于所有 } t$$

$Y(t)$和$X(t)$具有相同的有限阶分布，但显然以下两个事件的概率是不同的：

$$\{X(t) \leqslant 0.5 \ \text{对于所有} t\}$$

和

$$\{Y(t) \leqslant 0.5 \ \text{对于所有 } t\}$$

为了证明$Y(t)$和$X(t)$具有相同的n阶概率密度，需要找到$U=u$时X的n阶条件概率密度，然后对U积分求出结果。这个问题留给读者做练习。

例9.1-4的问题是互补事件$\{X(t)>0.5$ 对某些 $t \in [0,1]\}$中包括了不可数个随机变量。然而，统计测定和扩展的可加性公理4(参见8.1节)只允许我们对可数的随机变量的概率进行评估。接下来，通常假设随机过程“足够连续”，有限阶分布函数簇就足以确定这个随机过程所有时刻的值①。这种过程被称为“可分的”。上述例子中的随机过程$X(t)$显然是不可分的。

正如随机序列的情况，矩函数在实际应用中发挥了重要的作用。均值函数，用$\mu_X(t)$表示，定义为

$$\mu_X(t) \triangleq E[X(t)], \qquad -\infty < t < +\infty \tag{9.1-1}$$

类似地，相关函数定义为随机变量共轭积的均值

$$R_{XX}(t_1,t_2) \triangleq E[X(t_1)X^*(t_2)], \qquad -\infty < t_1,t_2 < +\infty \tag{9.1-2}$$

协方差函数定义为中心过程$X_c(t) \triangleq X(t)-\mu_X(t)$在$t_1$和$t_2$时刻共轭积的均值

$$\begin{aligned} K_{XX}(t_1,t_2) &\triangleq E[X_c(t_1)X_c^*(t_2)] \\ &\triangleq E[(X(t_1)-\mu_X(t_1))(X(t_2)-\mu_X(t_2))^*] \end{aligned} \tag{9.1-3}$$

显然这三个函数并不是没有关系的，事实上我们有

$$K_{XX}(t_1,t_2) = R_{XX}(t_1,t_2) - \mu_X(t_1)\mu_X^*(t_2) \tag{9.1-4}$$

我们一般也将方差定义为$\sigma_X^2(t) \triangleq K_{XX}(t,t) = E[|X_c(t)|^2]$，功率函数$R_{XX}(t,t)=E[|X(t)|^2]$。

例9.1-5 （再论随机正弦波）设随机过程

$$X(t) = A\sin(\omega_0 t + \Theta)$$

其中A和Θ是相互独立的实随机变量，Θ在$[-\pi,+\pi]$上服从均匀分布。对这个正弦随机过程，可以得到均值函数和相关函数分别为

$$\begin{aligned} \mu_X(t) &= E[A\sin(\omega_0 t + \Theta)] \\ &= E[A]E[\sin(\omega_0 t + \Theta)] \\ &= \mu_A \cdot \frac{1}{2\pi}\int_{-\pi}^{+\pi} \sin(\omega_0 t + \theta)\mathrm{d}\theta \\ &= \mu_A \cdot 0 = 0 \end{aligned}$$

---

① 9.3节介绍的白噪声是一个例外。

$$\begin{aligned}R_{XX}(t_1,t_2) &= E[X(t_1)X^*(t_2)]\\ &= E[A^2\sin(\omega_0 t_1+\Theta)\sin(\omega_0 t_2+\Theta)]\\ &= E[A^2]E[\sin(\omega_0 t_1+\Theta)\sin(\omega_0 t_2+\Theta)]\end{aligned}$$

利用三角恒等式

$$\sin(B)\sin(C)=\tfrac{1}{2}\{\cos(B-C)-\cos(B+C)\}$$

第二项可写为

$$\tfrac{1}{2}\{E[\cos(\omega_0(t_1-t_2))]-E[\cos(\omega_0(t_1+t_2)+2\Theta)]\} \tag{9.1-5}$$

由于Θ在$[-\pi,\ +\pi]$上均匀分布，式(9.1-5)中第二个数学期望为零，最终得到

$$R_{XX}(t_1,\ t_2)=\frac{1}{2}E[A^2]\cos\omega_0(t_1-t_2)$$

我们注意到$\mu_X(t)=0$(常数)，$R_{XX}(t_1,\ t_2)$只与时间差t_1-t_2有关，这类随机过程在9.4节中被称为广义平稳随机过程。

像离散时间信号那样，相关函数和协方差函数是共轭对称的，即

$$R_{XX}(t_1,t_2)=R^*_{XX}(t_2,t_1)$$

$$K_{XX}(t_1,t_2)=K^*_{XX}(t_2,t_1)$$

直接根据均值算子E的线性特性可得到。

如果对随机过程在t_1，t_2，…，t_N这N个时刻进行采样，则构造了一个随机向量。我们已经知道，一个随机向量的相关或协方差矩阵必须是半正定的(参见第5章)。当然这对随机过程的相关函数和协方差函数提出了一定的要求。具体来说，每一个由相关(协方差)函数得到的相关(协方差)矩阵必须是半正定的。接下来我们定义半正定函数。

定义9.1-3　如果对所有的$N>0$和$t_1<t_2<\cdots<t_N$以及所有的复常数a_1，a_2，…，a_N满足

$$\sum_{i=1}^{N}\sum_{j=1}^{N}a_i a_j^* g(t_i,t_j)\geqslant 0$$

则二维函数$g(t,\ s)$是半正定的。

运用这个定义，可以说，所有的相关和协方差函数必须是半正定的。稍后我们将看到，这个必要条件也是充分的。虽然半正定是一个重要的约束，但很难应用这个条件来对提出相关函数的合法性进行测试。

相关和协方差函数的另一个基本的属性是对角优势，即

$$|R_{XX}(t,s)|\leqslant\sqrt{R_{XX}(t,t)R_{XX}(s,s)},\ \text{对于所有}\ t,s$$

这可由柯西-施瓦茨不等式得到[参见式(4.3-17)]。半正定矩阵就意味着对角优势，但这是个较弱的条件。

9.2　一些重要的随机过程

在这一节中将介绍几个重要的随机过程。我们从异步二进制信号(ABS)和随机电报信号(RTS)开始，然后是泊松计数过程，相移键控(PSK)随机过程，一个数字调制信号实例以及由随机游动序列的极限所构造的维纳过程，最后介绍更广一类的马尔可夫过程。

异步二进制信号

图 9.2-1 是异步二进制信号(ABS)的一条样本函数(对数字调制和计算机非常重要)。每个脉冲的宽度为 T，随机变量 X_n 表示第 n 个脉冲的高度，等概率取 $\pm a$。

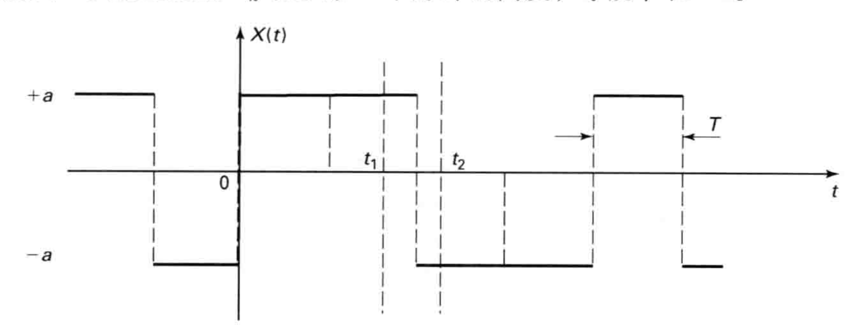


图 9.2-1 异步二进制信号(ABS)的一条样本函数(设 $D=0$)

序列是异步的，这是因为第 n 个脉冲的起始时间，或者第零个脉冲的位移 D 是在 $U\left(-\frac{T}{2},\frac{T}{2}\right)$ 上均匀分布的随机变量。对于 $|t_2-t_1|<T$，采样时刻 t_2 和采样时刻 t_1 可能在同一个脉冲内，也有可能在不同的脉冲上。

ABS 过程可以用数学模型描述为

$$X(t)=\sum_n X_n\, w\left[\frac{t-D-nT}{T}\right]$$

其中脉冲(矩形窗)函数 $w(t)$ 定义为

$$w(t)\triangleq\begin{cases}1, & |t|\leqslant\frac{1}{2}\\ 0, & \text{其他}\end{cases}$$

这个实过程的相关函数为

$$\begin{aligned}R_{XX}(t_1,t_2)&=E[X(t_1)X(t_2)]\\&=E\left[\sum_n\sum_l X_nX_l\, w\left(\frac{t_1-D-nT}{T}\right)\, w\left(\frac{t_2-D-lT}{T}\right)\right]\end{aligned}$$

在 ABS 过程中，假定不同脉冲的电平是独立的随机变量，并且独立于随机位移 D。由于 $n\neq l$ 时 $E[X_nX_l]=E[X_n]E[X_l]$，$E[X_n^2]=a^2$，于是得到

$$\begin{aligned}R_{XX}(t_1,t_2)=a^2\sum_n E\left[w\left(\frac{t_1-D-nT}{T}\right)\, w\left(\frac{t_2-D-nT}{T}\right)\right]\\+\sum_{n\neq l}\sum_l E[X_n]E[X_l]\, E\left[w\left(\frac{t_1-D-nT}{T}\right)\, w\left(\frac{t_2-D-lT}{T}\right)\right]\end{aligned}$$

式子右边的第二项包括 $n\neq l$ 的条件，因为 $E[X_n]=E[X_l]=0$，所以

$$\begin{aligned}&\sum_n E\left[w\left(\frac{t_1-D-nT}{T}\right)\, w\left(\frac{t_2-D-nT}{T}\right)\right]\\&=\sum_n\frac{1}{T}\int_{t_2-\frac{T}{2}-nT}^{t_2+\frac{T}{2}-nT} w\left(\frac{\alpha}{T}\right)\, w\left(\frac{\alpha-(t_2-t_1)}{T}\right)\mathrm{d}\alpha\\&=\left(1-\frac{(t_2-t_1)}{T}\right)\, w\left(\frac{t_2-t_1}{2T}\right),\qquad t_2>t_1\end{aligned}$$

更一般地，定义 $\tau \triangleq t_2 - t_1 \leqslant 0$，由于 $w(|\tau|) = w(\tau)$，可以得到

$$R_{XX}(\tau) = a^2 \left(1 - \frac{|\tau|}{T}\right) w\left(\frac{\tau}{2T}\right) \tag{9.2-1}$$

式(9.2-1)还可以直接推广到任意两个电平之间等概率转换的情况下，比如 a 和 b，此时需要的修改为

$$R_{XX}(\tau) = \frac{1}{4}(a-b)^2 \left(1 - \frac{|\tau|}{T}\right) w\left(\frac{\tau}{2T}\right) + \left(\frac{a+b}{2}\right)^2$$

我们将这个求解过程留给读者作为练习。图 9.2-2 中显示的是 $a=1$，$b=0$ 以及 $T=1$ 时 ABS 的自相关函数 $R_{XX}(\tau)$。

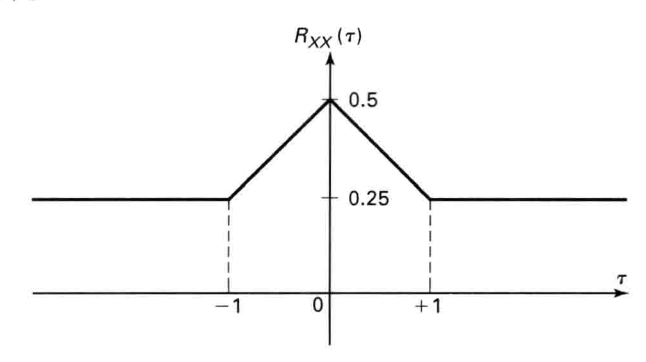


图 9.2-2 $a=1$，$b=0$ 以及 $T=1$ 时 ABS 的自相关函数 $R_{XX}(\tau)$

泊松计数过程

设过程 $N(t)$ 表示到时刻 t 所有计数(到达)的数目。则 $N(t)$ 可以写为

$$N(t) \triangleq \sum_{n=1}^{\infty} u(t - T[n])$$

其中 $u(t)$ 是单位阶跃函数，$T[n]$ 是第 n 个到达的时间，它是例 8.1-11 中提到的时间随机序列。我们已经证明了 $T[n]$ 服从非平稳的一阶 Erlang 密度

$$f_T(t;n) = \frac{(\lambda t)^{n-1}}{(n-1)!} \lambda \mathrm{e}^{-\lambda t} u(t), \qquad n \geqslant 0 \tag{9.2-2}$$

这是对指数概率密度的 n 重卷积得到的结果。图 9.2-3 给出了一个典型的样本函数，其中 $T[n] = t_n$，$\tau[n] = \tau_n$，注意各个到达之间的时间

$$\tau[n] \triangleq T[n] - T[n-1]$$

称为到达时距，它们是联合独立且同布的，且服从指数分布

$$f_{\tau}(t) = \lambda \mathrm{e}^{-\lambda t} u(t)$$

这在例 8.1-11 中已证明。因此如果从 $t=0$ 开始计数，则 $T[n]$就表示了直到第 n 次到达之前的所有时间。

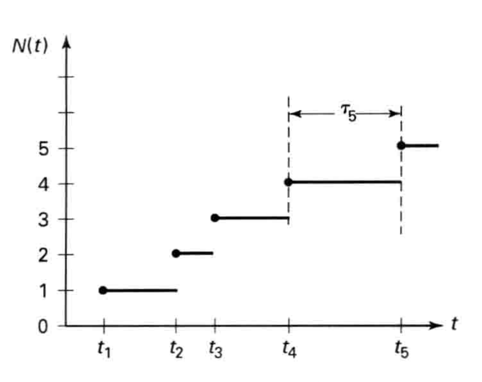


图 9.2-3 泊松过程的一条样本函数在 [0 , ∞) 上

$N(t)$是t时间以前(包括t时刻)总的到达数目，所以

$$P[N(t)=n]=P[T[n]\leqslant t,T[n+1]>t]$$

因为只有当随机变量$T[n]$小于或等于t以及随机变量$T[n+1]$大于t时，$N(t)$才能等于n。如果将独立的到达时距代入，则这个概率可表示为

$$P[T[n]\leqslant t,\tau[n+1]>t-T[n]]$$

运用到达时间$T[n]$与到达时距$\tau[n+1]$的统计独立性，很容易计算出如下结果：

$$\int_0^t f_T(\alpha;n)\left[\int_{t-\alpha}^{\infty}f_\tau(\beta)\mathrm{d}\beta\right]\mathrm{d}\alpha=\int_0^t\frac{\lambda^n\alpha^{n-1}\mathrm{e}^{-\lambda\alpha}}{(n-1)!}\left(\int_{t-\alpha}^{\infty}\lambda\mathrm{e}^{-\lambda\beta}\mathrm{d}\beta\right)\mathrm{d}\alpha\cdot u(t)$$
$$=\left(\int_0^t\alpha^{n-1}\mathrm{d}\alpha\right)\lambda^n\mathrm{e}^{-\lambda t}/(n-1)!\ u(t)$$

利用$P_N(n;t)\triangleq P\{N(t)=n\}$

$$P_N(n;t)=\frac{(\lambda t)^n}{n!}\mathrm{e}^{-\lambda t}u(t),\qquad t\geqslant0,\qquad n\geqslant0\tag{9.2-3}$$

这样就得到了泊松计数过程的 PMF。注意到，它与泊松随机变量(均值为$\mu=\lambda t$)的 PMF 相等[参见式(2.5-13)和式(1.10-5)]，则

$$E[N(t)]=\lambda t\tag{9.2-4}$$

我们称λ为平均到达率(有时也被称为强度)。这是符合大家的直观认识的，泊松过程在时间t的均值等于平均到达率λ乘以时间间隔$(0,t]$的长度，具体的分析留给读者课后练习。

由于随机序列$T[n]$具有独立的增量(参见定义 8.1-4)，泊松过程定义所用的单位阶跃函数是因果的，泊松过程$N(t)$也有独立的增量看起来是合理的。然而，这样的结果是不明确的，因为联合独立的到达时距$\tau[n]$可能分布在两个不相交区间，从而造成邻近增量的依赖。然而，利用指数 PDF 的无记忆性(参见习题 9.8)，我们可以证明泊松过程确实具有独立增量的特点。

利用独立增量，可以得到泊松计数过程在区间(t_a,t_b)增量的 PMF 为

$$P[N(t_b)-N(t_a)=n]=\frac{[\lambda(t_b-t_a)]^n}{n!}\mathrm{e}^{-\lambda(t_b-t_a)}u(n)\tag{9.2-5}$$

其中利用了时距序列是平稳的条件，也就是λ为一个常数。我们用下面的定义描述出来。

定义 9.2-1 一个随机过程具有独立增量，是指n个随机变量的集合

$$X(t_1),X(t_2)-X(t_1),\cdots,X(t_n)-X(t_{n-1})$$

对所有$t_1<t_2<\cdots<t_n$和所有$n\geqslant1$都满足联合独立性。

这就是说，当相应的区间不重叠时增量是统计独立的。正如在随机序列的情况那样，独立增量性使得高阶分布较容易得到。例如，在目前的情况那样，对泊松计数过程，假设$t_2>t_1$，我们可以写成

$$P_N(n_1,n_2;t_1,t_2)=P[N(t_1)=n_1]\ P[N(t_2)-N(t_1)=n_2-n_1]$$
$$=\frac{(\lambda t_1)^{n_1}}{n_1!}\mathrm{e}^{-\lambda t_1}\frac{[\lambda(t_2-t_1)]^{n_2-n_1}}{(n_2-n_1)!}\mathrm{e}^{-\lambda(t_2-t_1)}u(n_1)u(n_2-n_1)$$

化简后得到

$$P_N(n_1,n_2;t_1,t_2)=\frac{\lambda^{n_2}t_1^{n_1}(t_2-t_1)^{n_2-n_1}}{n_1!(n_2-n_1)!}\mathrm{e}^{-\lambda t_2}u(n_1)u(n_2-n_1),\qquad 0\leqslant t_1<t_2$$

参见习题 1.54。利用独立增量性可以给出泊松计数过程的另一个定义。

定义9.2-2 泊松计数过程就是一个增量服从泊松分布[参见式(9.2-5)]的独立增量过程。

考虑泊松过程的矩函数，一阶矩已被证明为λt，它是过程的均值函数。令$t_2 \geqslant t_1$，我们可以利用独立增量性计算相关函数为

$$\begin{aligned} E[N(t_2)N(t_1)] &= E[(N(t_1)+[N(t_2)-N(t_1)])N(t_1)] \\ &= E[N^2(t_1)]+E[N(t_2)-N(t_1)]E[N(t_1)] \\ &= \lambda t_1+\lambda^2 t_1^2+\lambda(t_2-t_1)\lambda t_1 \\ &= \lambda t_1+\lambda^2 t_1 t_2 \end{aligned}$$

如果$t_2<t_1$，我们只需将公式中的t_1和t_2互换即可。于是对任意t_1和t_2都有结论

$$\begin{aligned} R_{\mathrm{NN}}(t_1,t_2) &= E[N(t_1)N(t_2)] \\ &= \lambda \min(t_1,t_2)+\lambda^2 t_1 t_2 \end{aligned} \tag{9.2-6}$$

利用式(9.2-4)和式(9.2-6)，可计算协方差为

$$K_{\mathrm{NN}}(t_1,t_2) = \lambda \min(t_1,t_2) \tag{9.2-7}$$

因此我们看到，该过程的方差等于λt，与均值相同，这是从泊松随机变量继承的属性。我们还看到，协方差仅仅取决于所关注的两个时刻中较早的那个。这是因为可以将$N(t)$描述为一个较早的时刻值加上一个增量，注意到增量的独立性以及$N(t)$在较早的时间意味着它们之间的协方差必须为零。因此，这种独立增量过程的协方差只是过程的方差在两个时刻中较早的那个。

例9.2-1 （放射性物质监测）在放射性监测系统中，粒子计数过程往往可用泊松过程进行建模。设计数器在任意时间t开始计数，持续计数T_0秒。如果计数值超过一个阈值，假设为N_0，警报就会响起。假设到达率是λ，我们想知道，当放射性物质出现时报警器不响的概率大小。

因为这是个泊松过程，我们知道它有满足泊松分布的独立增量。因此，在区间$(t, t+T_0]$的计数值为ΔN，则$\Delta N \triangleq N(t+T_0)-N(t)$是服从泊松分布的，其均值为$\lambda T_0$，与$t$无关。计数值$\Delta N$小于等于$N_0$的概率为

$$P[\Delta N \leqslant N_0] = \sum_{k=0}^{N_0} \frac{(\lambda T_0)^k}{k!} \mathrm{e}^{-\lambda T_0}$$

如果N_0很小，我们可以直接计算这个求和值。如果$\lambda T_0 \gg 1$，可以对泊松分布使用高斯近似[参见式(1.11-9)]。

例9.2-2 （两个独立的泊松过程的和）设$N_1(t)$是一个到达率为λ_1的泊松计数过程。$N_2(t)$是到达率为λ_2的另一个泊松计数过程，N_2独立于N_1。这两个过程的和$N(t) \triangleq N_1(t)+N_2(t)$，可作为两个独立机器的故障总数的模型，两机器的故障率分别为λ_1和λ_2。显然$N(t)$也是一个到达率为$\lambda=\lambda_1+\lambda_2$的泊松计数过程。

为了说明这点，我们使用泊松计数过程的定义9.2-2并且验证$N(t)$的这些条件。首先，如果两个独立增量过程是联合独立的，则它们的和也是一独立增量过程。其次，对任意增量$N(t_b)-N(t_a)\,(t_b>t_a)$，可写成

$$N(t_b)-N(t_a) = N_1(t_b)-N_1(t_a)+N_2(t_b)-N_2(t_a)$$

因此，N 中的增量是 N_1 和 N_2 两个相应增量的和。则预期的结果就可由这样的事实得出：两个独立的泊松随机变量的总和也是服从泊松分布的，且其参数等于这两个参数的总和(参见例 3.3-8)。因此，$N(t)$ 中增量的参数为

$$\lambda_1(t_b - t_a) + \lambda_2(t_b - t_a) = (\lambda_1 + \lambda_2)(t_b - t_a)$$

泊松计数过程 $N(t)$ 可以用几种方法推广。我们可以令到达率是时间的函数。到达率 $\lambda(t)$ 必须满足 $\lambda(t) \geqslant 0$。那么所产生的非均匀泊松计数过程的平均值就变成了

$$\mu_X(t) = \int_0^t \lambda(\tau)\mathrm{d}\tau, \qquad t \geqslant 0 \tag{9.2-8}$$

增量于是变成了独立的泊松分布，增量的均值则由这个时变均值函数确定。另一个可能的推广方法是二维的或空间的泊松过程，用于为图像传感器的光子到达或半导体晶片上的缺陷等建模。

泊松过程的另一种推导

如果从 1.10 节中列出的时间上随机点的基本性质去重新推导泊松计数过程将是非常有趣的。这里我们采用本章中用到的符号重新推导。对于小的 Δt

(1) $P_N(1;\ t,\ t+\Delta t) = \lambda(t)\Delta t + o(\Delta t)$。

(2) $P_N(k;\ t,\ t+\Delta t) = o(\Delta t)$, $k > 1$。

(3) 在非重叠区间中的事件是统计独立的。

这里的符号 $o(\Delta t)$，读做“小 oh”，是指比线性速率更快趋于 0 的任何量，因此

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

$$P_N(k; t, t+\Delta t) = P[N(t+\Delta t) - N(t) = k]$$

我们注意到，性质(3)是记录发生在 $(0,\ t]$ 间事件数的计数过程 $N(t)$ 的独立增量特性。

我们计算在 $(t,\ t+\tau)$ 中事件发生 k 次的概率 $P_N(k;\ t,\ t+\tau)$ 如下：考虑 $P_N(k;\ t,\ t+\tau+\Delta t)$，如果 Δt 非常小，考虑到性质(1)和(2)，则要想在 $(t,\ t+\tau+\Delta t)$ 中事件发生 k 次，只有如下两种可能：

$$E_1 = \{k \text{ 在区间} (t, t+\tau) \ \ \text{和} \ \ 0 \text{ 在区间} (t+\tau, t+\tau+\Delta t)\}$$

$$E_2 = \{k-1 \text{ 在区间} (t, t+\tau) \ \ \text{和} \ \ 1 \text{ 在区间} (t+\tau, t+\tau+\Delta t)\}$$

由于 E_1 和 E_2 是不相交事件，将其概率相加，可得到

$$\begin{aligned} P_N(k; t, t+\tau+\Delta t) &= P_N(k; t, t+\tau)P_N(0; t+\tau, t+\tau+\Delta t) \\ &\quad + P_N(k-1; t, t+\tau)P_N(1; t+\tau, t+\tau+\Delta t) \\ &= P_N(k; t, t+\tau)[1 - \lambda(t+\tau)\Delta t] \\ &\quad + P_N(k-1; t, t+\tau)\lambda(t+\tau)\Delta t \end{aligned}$$

将上式整理，除以 Δt，并取极限，则得到线性微分方程(LDE)

$$\frac{\mathrm{d}P_N(k; t, t+\tau)}{\mathrm{d}\tau} = \lambda(t+\tau)[P_N(k-1; t, t+\tau) - P_N(k; t, t+\tau)]$$

这样我们就得到了一组递归的一阶微分方程，令 $P_N(-1;\ t,\ t+\tau) = 0$，因为这是一个不可能事件的概率，于是，由上述一阶微分方程可以解出 $P_N(k;\ t,\ t+\tau)$，$k = 0, 1, \cdots$。同时，为了简化符号，暂时定义 $P_N(k) \triangleq P_N(k;\ t,\ t+\tau)$，这样，对 t 和 τ 的依赖就隐去了，但依然存在。

当 $k=0$ 时

$$\frac{\mathrm{d}P_N(0)}{\mathrm{d}\tau} = -\lambda(t+\tau)P_N(0)$$

这是一个简单的一阶齐次微分方程，由于 $P_N(0;t,t)=1$，$C=1$ 和

$$P_N(0) = \exp\left[-\int_t^{t+\tau} \lambda(\xi)\mathrm{d}\xi\right]$$

求得解为

$$P_N(0) = C\exp\left[-\int_t^{t+\tau} \lambda(\xi)\mathrm{d}\xi\right]$$

我们用下式来定义 μ：

$$\mu \triangleq \int_t^{t+\tau} \lambda(\xi)\mathrm{d}\xi$$

则

$$P_N(0) = \mathrm{e}^{-\mu}$$

当 $k=1$ 时，这个微分方程为

$$\begin{aligned}\frac{\mathrm{d}P_N(1)}{\mathrm{d}\tau} + \lambda(t+\tau)P_N(1) &= \lambda(t+\tau)P_N(0) \\ &= \lambda(t+\tau)\mathrm{e}^{-\mu}\end{aligned} \tag{9.2-9}$$

这个一阶非齐次方程的解是齐次方程的通解加上特解。对于通解 P_h，我们从 $k=0$ 的情况已经知道

$$P_h = C_2\mathrm{e}^{-\mu}$$

对于特解 P_p，采用参数变分法，设

$$P_p = v(t+\tau)\mathrm{e}^{-\mu}$$

其中 $v(t+\tau)$ 是要求的差。将这个方程代入式(9.2-9)，很容易得到

$$P_p = \mu\mathrm{e}^{-\mu}$$

则完整的解为 $P_N(1)=P_h+P_p$。由于 $P_N(1;t,t)=0$，可以得到 $C_2=0$，于是

$$P_N(1) = \mu\mathrm{e}^{-\mu}$$

一般情况 一般情况下 LDE 为

$$\frac{\mathrm{d}P_N(k)}{\mathrm{d}\tau} + \lambda(t+\tau)P_N(k) = \lambda(t+\tau)P_N(k-1)$$

通过归纳，可以得到

$$P_N(k) = \frac{\mu^k}{k!}\mathrm{e}^{-\mu}, \qquad k=0,1,\cdots$$

这是关键的结果。回顾一下 μ 的定义，可以得到

$$P_N(k;t,t+\tau) = \frac{1}{k!}\left[\int_t^{t+\tau} \lambda(\xi)\mathrm{d}\xi\right]^k \exp\left[-\int_t^{t+\tau} \lambda(\xi)\mathrm{d}\xi\right] \tag{9.2-10}$$

于是我们得到了非均匀的泊松计数过程。

另一种推广泊松过程的方法是对独立的到达时距使用不同的 pdf。一个更一般的过程，被称为“更新过程”[9-2]采用的就是非指数分布。词语“更新”即是将到达时间与设备的故障时间联系起来；因此计数过程 $N(t)$ 的值模拟了到目前为止必须补足的更新数量。

随机电报信号

当一个随机波形中所有的信息都包含在零交叉点中，常用所谓的“硬限幅器”来生成一个更简单但是等价的双电平波形，这样就与随机幅度的变化无关。一个特殊的情况是，当在一个时间间隔内过零点数服从泊松分布时，由此产生的随机过程称为随机电报信号(RTS)。图 9.2-4 显示的是 RTS 的一条样本函数。

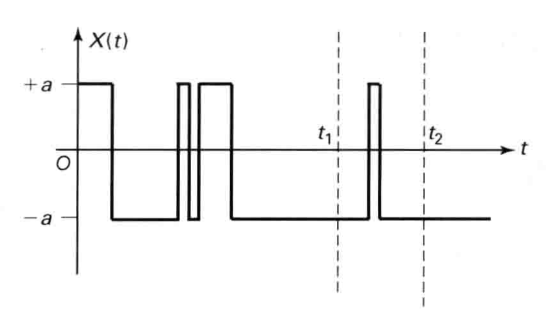


图 9.2-4 随机电报信号的样本函数

我们构建 $t \geqslant 0$ 的 RTS 如下：令 $X(0)$ 以等概率取 $\pm a$。然后以第 8 章所提到的泊松到达时间序列 $T[n]$ 的值来切换 RTS 的电平；即在 $T[1]$ 时刻改变 $X(t)$ 的符号，然后在 $T[2]$ 时刻又改变 $X(t)$ 的符号，然后一直下去。很显然由对称性以及到达时距 $\tau[n]$ 的平稳性，形成一个独立的随机序列，其均值 $\mu_X(t)=0$，一阶 PMF $P_X(a)=P_X(-a)=1/2$。接下来令 $t_2>t_1>0$，并考虑二阶 PMF $P_X(x_1, x_2) \triangleq P[X(t_1)=x_1, X(t_2)=x_2]$ 以及 $P_X(x_2|x_1) \triangleq P[X(t_2)=x_2|X(t_1)=x_1]$。由于 $P_X(a)=P_X(-a)=1/2$，我们可将相关函数写成

$$
\begin{aligned}
R_{XX}(t_1,t_2) &= E[X(t_1)X(t_2)] \\
&= a^2P_X(a,a)+(-a)^2P_X(-a,-a)+a(-a)P_X(a,-a)-a(a)P_X(-a,a) \\
&= \frac{1}{2}a^2(P_X(a|a)+P_X(-a|-a)-P_X(-a|a)-P_X(a|-a))
\end{aligned}
$$

但是 $P_X(-a|-a)=P_X(a|a)$ 刚好是在时距 $(t_1, t_2]$ 过零点为偶数的概率，而 $P_X(-a|a)=P_X(a|-a)$ 是过零点为奇数的概率。因此，记每单位时间转换的平均数为 λ，令 $\tau \triangleq t_2-t_1$，可得到

$$
R_{XX}(t_1,t_2)=a^2\left(\sum_{\text{偶数 } k\geqslant 0}\mathrm{e}^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}-\sum_{\text{奇数 } k\geqslant 0}\mathrm{e}^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}\right)=a^2\mathrm{e}^{-\lambda\tau}\sum_{\text{所有 } k\geqslant 0}(-1)^k\frac{(\lambda\tau)^k}{k!}
$$

其中用函数 $(-1)^k$ 将两个和联系起来，因为当 k 为偶数时，$(-1)^k=1$，当 k 为奇数时，$(-1)^k=-1$，因此，当 $\tau>0$ 时，我们有

$$
R_{XX}(t_1,t_2)=a^2\mathrm{e}^{-\lambda\tau}\sum_{\text{所有 } k\geqslant 0}\frac{(-\lambda\tau)^k}{k!}=a^2\mathrm{e}^{-2\lambda\tau}
$$

由于实过程的相关函数是对称的，即 $R_{XX}(t_1, t_2)=R_{XX}(t_2, t_1)$，所以当 $\tau \leqslant 0$ 时，可以将 $-\tau$ 代入上面的方程中，从而得到 $R_{XX}(t_1, t_2)=a^2\mathrm{e}^{+2\lambda\tau}$，于是，可得到对所有的间隔长度 τ

$$
R_{XX}(t_1,t_2)=a^2\mathrm{e}^{-2\lambda|\tau|}
$$

图 9.2-5 显示的是这个相关函数的图形。

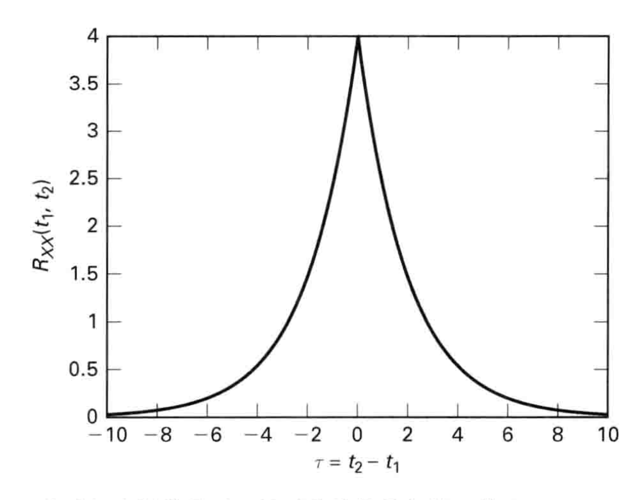


图 9.2-5　一个随机电报信号过程的对称的指数相关函数（$a = 2.0$，$\lambda = 0.25$）

相移键控数字调制

数字计算机产生许多二进制序列(数据)，这些序列被传递到其他数字计算机。这通常涉及某种形式的调制。二元调制方法可将这些数据频移到非常适合于传输介质(如电话线)的电磁频谱上。用于调制数据的一个基本方法是相移键控(PSK)。在这种方法中，由随机序列 $B[n]$ 建模的二元数据被逐位映射成相位序列 $\Theta[n]$，该相位序列用来调制载波信号 $\cos(2\pi f_c t)$。

特别地，当 $B[n]$ 是伯努利随机序列，并以等概率取 0 和 1 时，我们定义随机相位序列 $\Theta[n]$ 为

$$\Theta[n] \triangleq \begin{cases} +\pi/2, & B[n]=1 \\ -\pi/2, & B[n]=0 \end{cases}$$

用 $\Theta_a(t)$ 来表示模拟角过程，我们定义

$$\Theta_a(t) \triangleq \Theta[k], \qquad kT \leqslant t < (k+1)T$$

构造调制信号

$$X(t) = \cos(2\pi f_c t + \Theta_a(t)) \tag{9.2-11}$$

其中 T 为一个比特数据的传输持续时间。通常，T 选为 $1/f_c$ 的倍数，这样每比特时间 T 内就有整数个载波周期，T 的倒数称为消息或波特率。图 9.2-6 是整个调制器的示意图。该过程 $X(t)$ 是 PSK 过程。

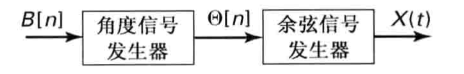


图 9.2-6　用伯努利随机序列 $B[n]$ 实现的 PSK 调制系统

我们的目标是计算随机 PSK 过程的均值函数和相关函数。为了便于计算，我们定义两个基函数

$$s_I(t) \triangleq \begin{cases} \cos(2\pi f_c t), & 0 \leqslant t \leqslant T \\ 0, & \text{其他} \end{cases}$$

和

$$s_Q(t) \triangleq \begin{cases} \sin(2\pi f_c t), & 0 \leqslant t \leqslant T \\ 0, & \text{其他} \end{cases}$$

则式(9.2-11)可以写成

$$\cos[2\pi f_c t+\Theta_a(t)] = \cos(\Theta_a(t))\cos 2\pi f_c t - \sin(\Theta_a(t))\sin 2\pi f_c t$$
$$= \sum_{k=-\infty}^{+\infty}\cos(\Theta[k])s_I(t-kT) - \sum_{k=-\infty}^{+\infty}\sin(\Theta[k])s_Q(t-kT) \quad (9.2\text{-}12)$$

$X(t)$ 的均值可由 $\cos(\Theta[n])$ 和 $\sin(\Theta[n])$ 的均值得到。由 $\Theta[n]$ 的定义，在 $\cos(\Theta[n])=0$，$\sin(\Theta[n])$ 以等概率取 ± 1 这种特殊情况下，可得到 $X(t)$ 的均值为 0，即 $\mu_X(t)=0$。

利用式(9.2-12)，可以计算相关函数为

$$R_{XX}(t_1,t_2) = \sum_{k,l} E\{\sin\Theta[k]\sin\Theta[l]\}s_Q(t_1-kT)s_Q(t_2-lT)$$

其中包括了随机序列 $\sin(\Theta[n])$ 的自相关函数

$$R_{\sin\Theta,\sin\Theta}[k,l] = \delta[k-l]$$

于是，$X(t)$ 的自相关函数变成了

$$R_{XX}(t_1,t_2) = \sum_{k=-\infty}^{+\infty} s_Q(t_1-kT)s_Q(t_2-kT) \quad (9.2\text{-}13)$$

由于 s_Q 只在 $[0,T]$ 中有值，所以，在式(9.2-13)中乘积项之间在 (t_1, t_2) 上没有重叠的项为零，在和中只有一个乘积项非零。此外，如果 t_1, t_2 不是同一周期，则自相关函数也为零。利用符号

$$((t)) \triangleq t \bmod T \quad 和 \quad \lfloor t/T \rfloor \triangleq 整数部分\,(t/T)$$

可以得到

$$R_{XX}(t_1,t_2) = \begin{cases} s_Q((t_1))s_Q((t_2)), & \lfloor t_1/T\rfloor = \lfloor t_2/T\rfloor \\ 0, & 其他 \end{cases}$$

特别是当 $0\leqslant t_1\leqslant T$, $0\leqslant t_2\leqslant T$ 时，我们有

$$R_{XX}(t_1,t_2) = s_Q(t_1)s_Q(t_2)$$

维纳过程或布朗运动

在例 8.1-13 中介绍过一个随机游动序列 $X[n]$。这里，我们构建一个类似的随机过程，它为分段常数过程，每段的长度为 T，这个过程表示为

$$X_T(t) \triangleq \sum_{k=1}^{\infty} W[k]u(t-kT)$$

其中

$$W[k] \triangleq \begin{cases} +s, & p=0.5 \\ -s, & p=0.5 \end{cases}$$

$u(t)$ 是连续的单位阶跃函数。

由于

$$X_T(nT) = \sum_{k=1}^{n} W[k] = X[n]$$

$X_T(nT)=X[n]$ 是随机游动序列。这样，就可以运用随机游动序列的已知结果计算这个过程的 PMF 和矩。维纳过程(或被称为维纳-列维或布朗运动)就是这样一个过程，其分布就是上

面所提到的分段常数过程的分布在间隔 T 趋于零的极限形式。令阶跃高度 s 以及时间间隔都趋于零，就获得了一个在极限条件下连续的随机过程，也就是，样本函数是时间连续函数的这样一个过程。在令 s 和 T 趋于零时，必须小心确保方差为有限的非零值。由此产生的维纳过程将继承独立增量特性。

维纳过程的最初建立是为气体分子的混乱随机运动进行建模的。用随机游动为基本离散碰撞建模，在一个无穷小的(非常小)的时间尺度内无限多(非常大)的分子相互作用，因此得到一个渐近过程。

在例8.1-13中，设 n 为试验次数，k 是成功的次数，$n-k$ 为失败的次数。$r \triangleq k-(n-k)=2k-n$ 表示成功次数超过失败次数的多少。那么 $2k=n+r$ 或 $k=(n+r)/2$，且必须是整数；不可能有2.5次成功。这样，$n+r$ 必须是偶数，$X_T(nT)=rs$ 的概率是在总共 n 次试验中，有 $0.5(n+r)$ 次成功($+s$)和 $0.5(n-r)$ 次失败($-s$)的概率，于是根据二项式的PMF，得到

$$P[X_T(nT)=rs]=\binom{n}{\frac{n+r}{2}}2^{-n},\ \text{当}\ n+r\ \text{为偶数时}$$

如果 $n+r$ 为奇数，则 $X_T(nT)$ 不可能等于 rs。

只要注意到随机变量 $X[n]$ 是8.1节定义的 n 个独立的伯努利随机变量的和，则均值、方差的计算就非常容易了。因此

$$E[X_T(nT)]=0$$

和

$$E[X_T^2(nT)]=ns^2$$

用 $t=nT$ 代入方差的表达式中，得到

$$\mathrm{Var}[X_T(t)]=E[X_T^2(nT)]=t\frac{s^2}{T}$$

这样就需要 s^2 与 T 成比例来得到一个有趣的极限分布①。我们设 $s^2=\alpha T$，其中 $\alpha>0$。随着 T 趋近于0我们保持方差始终为 αT。由中心极限定理的一个基本应用(参见4.7节)，可得到的极限为高斯分布。由于上述随机游动过程对所有的 T(无论多小)都有独立的增量，所以取极限得到的随机过程(在分布意义下收敛)为独立增量过程。因此，可得出以下极限过程的特性，这个过程称为维纳过程，其特性为

$$\mu_X(t)=0,\qquad \mathrm{Var}[X(t)]=\alpha t$$

$$f_X(x;t)=\frac{1}{\sqrt{2\pi\alpha t}}\exp\left(-\frac{x^2}{2\alpha t}\right),\qquad t>0 \tag{9.2-14}$$

所以对所有 $t>t$，增量 $\Delta \triangleq X(t)-X(\tau)$ 的PDF为

$$f_\Delta(\delta;t-\tau)=\frac{1}{\sqrt{2\pi\alpha(t-\tau)}}\exp\left(-\frac{\delta^2}{2\alpha(t-\tau)}\right) \tag{9.2-15}$$

这是由于

$$E[X(t)-X(\tau)]=E[\Delta]=0 \tag{9.2-16}$$

和

$$E\left[(X(t)-X(\tau))^2\right]=\alpha(t-\tau),\qquad t>\tau \tag{9.2-17}$$

---

① s^2 与 T 成比例的物理含义是：如果设置颗粒的速度是 $v \triangleq s/T$，为了保持两个量的乘积为一个常数，当位移量趋于零时，颗粒速度趋于无穷大。

例 9.2-3　(样本函数)我们可以使用 MATLAB 来对维纳过程的典型的样本函数进行可视化研究。因为它是计算机仿真的，对于固定的 $\alpha>0$，我们也可以对 $s=\sqrt{\alpha T}$趋近于 0 时对极限序列的影响进行评估。

从 1000 个元素的向量开始，它是伯努利随机向量 W 的一个现实，其中 $p=0.5$，产生过程如下：

```
u = rand(1000,1)
w = 0.5 >= u
```

下一行是对预先设定的 s，将 w 的范围转换到 $\pm s$

```
w = s*(2*w - 1.0)
```

然后产生一段样本函数 $X_T(nT)=X[n]$ 作为随机向量的元素

```
x = cumsum(w)
```

在数值实验中，令 $\alpha=1$，$T=0.01(s=0.1)$。对于 $T=0.01$，使用一个 1000 维的计算机变量，得到的结果 x 显示在图 9.2-7 中。尤其值得注意的是，在这种接近极限的情况下，方差随时间的增加所带来的影响。还要注意到随着时间的继续，这个趋势仍在发展。

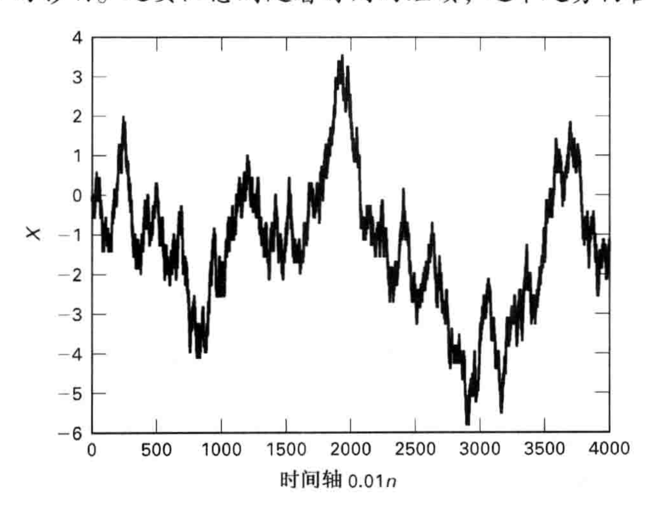


图 9.2-7　一个维纳过程的样本函数近似，$\alpha=1$，$T=0.01$

从 X 的一阶概率密度函数(pdf)和增量 Δ 的密度，有可能计算出 n 阶 pdf 的一个完备的集合，正如我们前面所看到的。因此得出，维纳过程的所有的 n 阶 pdf 是高斯的。

定义 9.2-3　如果对于所有的正整数 n，一个随机过程的 n 阶 pdf 都是联合高斯的，那么这个过程被称为高斯随机过程。

因此，维纳过程是高斯随机过程的一个例子。维纳过程的协方差函数(也就是它的相关函数，因为 $\mu_X(t)=0$)为

$$K_{XX}(t_1,t_2)=\alpha\min(t_1,t_2),\quad \alpha>0 \tag{9.2-18}$$

为了说明这点，假设 $t_1\geqslant t_2$，并且注意到增量 $X(t_1)-X(t_2)$ 与 $X(t_2)$ 独立，它们的均值都为 0

$$\begin{aligned}E[(X(t_1)-X(t_2))X(t_2)]&=E[X(t_1)-X(t_2)]E[X(t_2)]\\&=0\end{aligned}$$

或者

$$E[X(t_1)X(t_2)] = E[X^2(t_2)]$$
$$= \alpha t_2$$

如果 $t_2 > t_1$，我们得到 $E[X(t_2)X(t_1)] = \alpha t_1$，于是得到式(9.2-18)。

我们注意到，维纳过程具有和泊松过程相同的方差函数，尽管这两个过程是完全不同的。泊松过程仅是由常数值隔开的跳变组成。维纳过程则没有跳变，并且可以证明是几乎处处连续的；即样本函数依概率1连续。稍后，我们将证明，维纳过程是弱均方连续的(在网上的第10章中有更精确的证明)。

马尔可夫随机过程

到目前为止我们讨论了五种随机过程。其中，维纳和泊松过程是基本过程，许多其他更一般的随机过程已被证明是这两个基本过程的非线性变换得到的。在这两种情况下，通过运用独立增量特性解决了有联系的过程 n 阶统计特性描述的难题。事实上，这是一个相当通用的方法，开始我们得到任意一阶的分布，然后指定增量的分布，从而获得一组 n 阶分布来展示它们的联系。

另一种被证明是非常有用的，由一阶概率得到 n 阶概率的方法是马尔可夫过程方法。我们从一阶密度(或PMF)和条件密度(或条件PMF)开始

$$f_X(x;t) \quad \text{和} \quad f_X(x_2|x_1;t_2,t_1), \qquad t_2 > t_1$$

然后得到 n 阶 pdf $f(x_1, \cdots, x_n; t_1, \cdots, t_n)$ 如下：

$$f(x_1;t)f(x_2|x_1;t_2,t_1)\cdots f(x_n|x_{n-1};t_n,t_{n-1}) \tag{9.2-19}$$

我们要求读者证明，这是一个有效的 n 阶 pdf(即这个函数是非负的，积分为1)，只要条件和一阶 pdf 定义良好。

相反，如果从一个任意的 n 阶 pdf 开始，重复运用条件概率的定义可以得到

$$f(x_1,\cdots,x_n;t_1,\cdots,t_n) = f(x_1;t_1)f(x_2|x_1;t_2,t_1)f(x_3|x_2,x_1;t_3,t_2,t_1) \times \cdots \times f(x_n|x_{n-1},\cdots,x_1;t_n,\cdots,t_1) \tag{9.2-20}$$

如果约束条件概率密度仅取决于最近的条件值，则式(9.2-20)与式(9.2-19)相等。于是引发了下文对马尔可夫随机过程的定义。

定义 9.2-4 (马尔可夫随机过程)

(a)连续的(一阶)马尔可夫过程 $X(t)$ 满足条件 pdf

$$f_X(x_n|x_{n-1},x_{n-2},\cdots,x_1;t_n,\cdots,t_1) = f_X(x_n|x_{n-1};t_n,t_{n-1})$$

对所有 $x_1, x_2, \cdots, x_n$，$t_1 < t_2 < \cdots < t_n$ 以及所有整数 $n > 0$ 都成立。

(b)离散的(一阶)马尔可夫过程满足条件 PMF

$$P_X(x_n|x_{n-1},\cdots,x_1;t_n,\cdots,t_1) = P_X(x_n|x_{n-1};t_n,t_{n-1})$$

对所有 $x_1, \cdots, x_n$，$t_1 < \cdots < t_n$ 以及所有整数 $n > 0$ 都成立。

则 $X(t)$ 在某给定时刻 t 的值决定了这个过程未来值的条件概率。这个过程的值称为过程的状态，条件概率被认为是状态之间的转移概率。如果只有有限或可数集的值 x_i 是允许的，则离散的马尔可夫过程被称为马尔可夫链。马尔可夫链的一个例子是我们之前研究过的泊松

计数过程。维纳过程是连续马尔可夫过程的一个例子。因为它们的独立增量性，所以这两个过程都是马尔可夫过程。事实上，任何独立增量过程都是马尔可夫过程。为了说明这点，我们注意到，对离散值的情况，例如

$$
\begin{aligned}
&P_X(x_n|x_{n-1},\cdots,x_1;t_n,\cdots,t_1)\\
&\quad=P[X(t_n)=x_n|X(t_{n-1})=x_{n-1},\cdots,X(t_1)=x_1]\\
&\quad=P[X(t_n)-X(t_{n-1})=x_n-x_{n-1}|X(t_{n-1})=x_{n-1},\cdots,X(t_1)=x_1]\\
&\quad=P[X(t_n)-X(t_{n-1})=x_n-x_{n-1}]\quad\text{由独立增量性}\\
&\quad=P[X(t_n)-X(t_{n-1})=x_n-x_{n-1}|X(t_{n-1})=x_{n-1}]\qquad\text{再由独立增量性}\\
&\quad=P[X(t_n)=x_n|X(t_{n-1})=x_{n-1}]\\
&\quad=P_X(x_n|x_{n-1};t_n,t_1)
\end{aligned}
$$

然而要注意的是反过来结论是不成立的。一个马尔可夫随机过程不一定有独立的增量(参见习题9.17)。

马尔可夫随机过程在许多领域都得到了应用，包括信号处理，通信和控制系统。马尔可夫链应用在通信、计算机网络、可靠性理论等领域。

例 9.2-4　(多处理器的可靠性)给定一个有两个独立处理器的计算机，我们可以把它看成是一个具有三个状态的系统：0——两个处理器不工作；1——只有一个处理器工作，2——两个处理器都工作。我们想知道这三种状态的概率。一个常见的概率模型是处理器出现故障的时间为随机的，故障时间服从参数 $\lambda>0$ 的指数分布。一旦一个处理器出现故障，修复它的时间假定同为参数 $\mu>0$ 的指数分布。而且假定处理器的故障和维修时间是独立的；因此，在我们的概率模型中，故障和维修时间是联合独立的。

如果定义 $X(t)$ 为 t 时刻的系统状态，则 X 是一个连续时间的马尔可夫链。关于这一点，我们可以这样证明。首先证明 X 状态转移的时间是服从指数分布的，然后利用指数分布的无记忆性(参见习题9.8)。分析转移时间(故障时间或维修时间)可以继续做如下描述：从状态 $X=0$ 到 $X=1$ 的转移时间是两个指数分布维修时间的最小值，假定它们是相互独立的。在习题3.26中已经得到这个时间也服从参数为 2μ 的指数分布。则转移时间期望值为 $1/(2\mu)=\frac{1}{2}(1/\mu)$，也就是，维修一个处理器的平均时间的一半。这是合理的结论，因为状态 $X=0$ 时两个处理器都发生故障，因此需要同时、独立地接受维修，则从状态0过渡到状态1的率参数是 2μ。从状态1过渡到状态2要等待一个速率为 μ 的指数维修时间。也就是，它的速率也为 μ。同样，状态1转移到状态0只需等待一个率参数为 λ 的故障时间，而状态2到状态1等待两个指数分布的故障时间的最小值。因此它的率参数是 2λ。同时从状态0转移到状态2和状态2到状态0的概率是0，因此可以忽略。

图9.2-8显示了这个马尔可夫链模型的短时状态转移图。图中的有向支路表示的是短时，即当 $\Delta t\to 0$ 时，状态之间的转移概率。转移时间假设服从指数分布，分布参数则在分支标签上给出。这些转移时间称为内部转移时间可能更恰当，类似于泊松计数过程的到达时距，它同样服从指数分布。

考虑 $t+\Delta t$ 时刻处于状态2以及在 t 时刻已处于状态1的概率，则需要维修时间 $T_s(T_s\geqslant t)$ 处于间隔 $(t,t+\Delta t]$ 之间，令 $P_i(t)\triangleq P[X(t)=i]$，其中 $0\leqslant i\leqslant 2$，则

$$P_2(t+\Delta t)=P_1(t)P[t<T_s\leqslant t+\Delta t|T_s\geqslant t]$$

其中

$$P[t<T_s\leqslant t+\Delta t|T_s\geqslant t]=\frac{F_{T_s}(t+\Delta t)-F_{T_s}(t)}{1-F_{T_s}(t)}=\mu\Delta t+o(\Delta t)$$

利用这类变量来连接从 t 时刻的状态到 $t+\Delta t$ 时刻的转移概率，忽略从状态2到状态0的转移，使得我们能够根据 t 时刻的状态概率写出 $t+\Delta t$ 时刻的状态概率，其向量形式为

$$\begin{bmatrix}P_0(t+\Delta t)\\P_1(t+\Delta t)\\P_2(t+\Delta t)\end{bmatrix}=\begin{bmatrix}1-2\mu\,\Delta t & \lambda\,\Delta t & 0\\ 2\mu\,\Delta t & 1-(\lambda+\mu)\,\Delta t & 2\lambda\,\Delta t\\ 0 & \mu\,\Delta t & 1-2\lambda\,\Delta t\end{bmatrix}\begin{bmatrix}P_0(t)\\P_1(t)\\P_2(t)\end{bmatrix}+\boldsymbol{o}(\Delta t)$$

其中 $\boldsymbol{o}(\Delta t)$ 表示的是比 Δt 低阶的微小量。

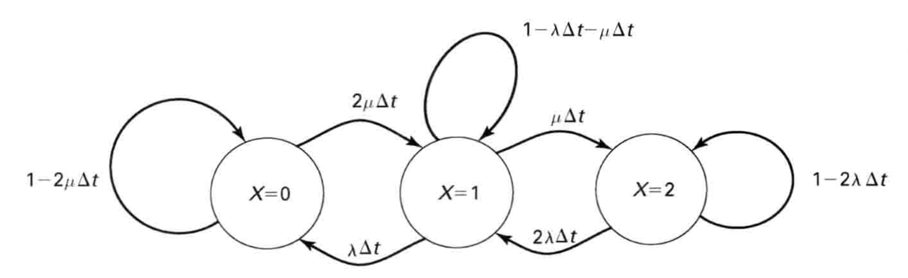


图9.2-8 标明了转移概率的短时状态转移图

整理后得到

$$\begin{bmatrix}P_0(t+\Delta t)-P_0(t)\\P_1(t+\Delta t)-P_1(t)\\P_2(t+\Delta t)-P_2(t)\end{bmatrix}=\begin{bmatrix}-2\mu & \lambda & 0\\ 2\mu & -(\lambda+\mu) & 2\lambda\\ 0 & \mu & -2\lambda\end{bmatrix}\begin{bmatrix}P_0(t)\\P_1(t)\\P_2(t)\end{bmatrix}\Delta t+\boldsymbol{o}(\Delta t)$$

两边都除以 Δt，运用矩阵的表示，可得到

$$\frac{\mathrm{d}\boldsymbol{P}(t)}{\mathrm{d}t}=\boldsymbol{AP}(t) \tag{9.2-21}$$

矩阵 $\boldsymbol{A}$ 称为马尔可夫链 X 的生成器。这个一阶向量微分方程可以用初始概率向量 $\boldsymbol{P}(0)\triangleq\boldsymbol{P}_0$，运用线性系统理论[9-3]求解出来。其解以指数向量表示如下：

$$\mathrm{e}^{\boldsymbol{A}t}\triangleq\boldsymbol{I}+\boldsymbol{A}t+\frac{1}{2!}(\boldsymbol{A}t)^2+\frac{1}{3!}(\boldsymbol{A}t)^3+\cdots$$

对所有有限的 t，上式都是收敛的。那么解 $\boldsymbol{P}(t)$ 为

$$\boldsymbol{P}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{P}_0,\qquad t\geqslant 0$$

这种方法的详细过程以及如何得到精确的解参见参考文献[9-4]。

目前，我们通过将式(9.2-21)关于时间的导数置为0，可得到稳态解，此时 $\boldsymbol{AP}=\boldsymbol{0}$。从第一行和最后一行，可得到

$$-2\mu P_0+\lambda P_1=0$$

以及

$$+\mu P_1-2\lambda P_2=0$$

于是我们得到 $P_1=(2\mu/\lambda)P_0$ 和 $P_2=(\mu/2\lambda)P_1=(\mu/\lambda)^2P_0$。然后利用 $P_0+P_1+P_2=1$，可得到 $P_0=\lambda^2/(\lambda^2+2\mu\lambda+\mu^2)$，最后得到

$$\boldsymbol{P}=\frac{1}{\lambda^2+2\mu\lambda+\mu^2}[\lambda^2,2\mu\lambda,\mu^2]^{\mathrm{T}}$$

这样两个处理器都出故障的稳态概率为 $P_0=[\lambda/(\lambda+\mu)]^2$。而如果我们只用了一个由两状态马尔可夫链建模的处理器，得到的结果则为 $P_0=\lambda/(\lambda+\mu)$。

显然我们可以将这个例子推广应用到任意 n 个状态的情况，只要这些状态之间的时间间隔是独立的且服从指数分布。事实上，这一过程称为一个排队过程。其他的例子如高速公路收费站的排队数和计算机中或电话网络的拥塞状态。更多关于排队系统的问题参见参考文献[9-2]。最后一个例子中非常重要的一点是呈指数分布的转移时间在体现马尔可夫性时是非常关键的。事实上，除了指数分布之外的任何其他分布都是无记忆的，由此产生的状态转移过程将不会是一个马尔可夫链。

生-死马尔可夫链

一个只允许在相邻状态中转换的马尔可夫链被称为生-死链。我们首先研究无限状态的情况，然后再研究有限状态的情况。

1. 无限长队列 图 9.2-9① 显示的是无限长队列的状态转移图。从状态 i 到状态 $i+1$，我们说"一次出生"发生了。同样，从状态 i 到状态 $i-1$，我们说"一次死亡"发生了。在任何时间 t，$P_j(t)$是处在状态 j 的概率，即有"人口" j，换句话说，即出生超过死亡的数目。在这个模型中，出生是由泊松过程产生的。出生之间的间隔用 τ_B 表示，死亡之间的间隔用 τ_D 表示，取决于状态而且分别服从参数为 λ_i 和 μ_i 指数分布。该模型被广泛应用在排队论中，其中一次出生是一次到达队列，一次死亡是一次从队列中离开。一个例子是排队等候的人在一个单一的售票处买票。如果剧院非常大而且对排队的长度没有限制(例如，队伍可能阻塞了人行道并带来危险)，即溢出和饱和先不考虑。则队列的动态变化可用基本方程 $W_n=\max\{0,W_{n-1}+\tau_s-\tau_i\}$ 来描述，其中 W_n 是第 n 个到达的人在队列中的等待时间，τ_s 是为第$(n-1)$次到达的人的服务时间，τ_i 是第 n 个和第$(n-1)$个到达的人之间的间隔时间。这是不受限制的队列长度的一个例子。另一方面，存储在一个有限大小的缓冲区的数据包问题则是一个不同的问题。当缓冲区被填满(饱和)时，一个新的数据到来必然被拒绝(在这种情况下，我们说数据包"丢失")。

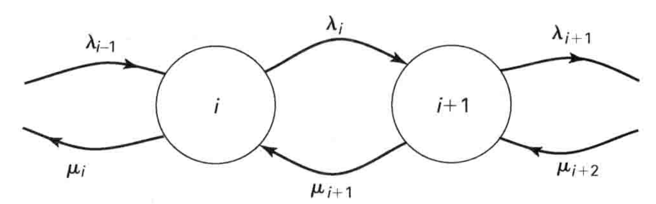


图 9.2-9 标出了转移率参数的生-死马尔可夫状态图

继续例 9.2-4 中的过程，我们可以写为

$$\boldsymbol{P}(t+\Delta t)=\boldsymbol{BP}(t)$$

① 我们画的图只显示了转移率参数，即在状态之间链接上标注的 μ_i 和 λ_i。这种图显示的不是很清晰，例如，在泊松过程中处在状态 i 的短时概率是 $1-(\lambda_i+\mu_i)\Delta t$，尽管这种图不是很清晰，但它显的不那么挤，图 9.2-8 是非标准短时转移概率图。

其中

$$
\boldsymbol{B}=\begin{bmatrix}
1-\lambda_0\Delta t & \mu_1\Delta t & 0 & \cdots & \\
\lambda_0\Delta t & 1-(\lambda_1+\mu_1)\Delta t & \mu_2\Delta t & 0 & \cdots \\
0 & \lambda_1\Delta t & 1-(\lambda_2+\mu_2)\Delta t & \mu_3\Delta t & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots
\end{bmatrix}
$$

重新整理，除以 Δt，并且令 $\Delta t\to 0$，可得到

$$
\mathrm{d}\boldsymbol{P}(t)/\mathrm{d}t=\boldsymbol{A}\boldsymbol{P}(t)
$$

其中 $\boldsymbol{P}(t)=[P_0(t),P_1(t),\cdots,P_j(t),\cdots]^{\mathrm{T}}$，马尔可夫链的产生矩阵 $\boldsymbol{A}$ 为

$$
\boldsymbol{A}=\begin{bmatrix}
-\lambda_0 & \mu_1 & 0 & \cdots & \\
\lambda_0 & -(\lambda_1+\mu_1) & \mu_2 & 0 & \cdots \\
0 & \lambda_1 & -(\lambda_2+\mu_2) & \mu_2 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots
\end{bmatrix}
$$

在稳态时 $\boldsymbol{P}'(t)=\boldsymbol{0}$，这样从 $\boldsymbol{AP}=\boldsymbol{0}$ 就可以得到

$$
\begin{aligned}
P_1&=\rho_1P_0\\
P_2&=\rho_2P_1=\rho_1\rho_2P_0\\
&\vdots\\
P_j&=\rho_jP_{j-1}=\rho_j\cdots\rho_2\rho_1P_0
\end{aligned}
$$

其中当 $j\geqslant 1$ 时 $\rho_j\triangleq\lambda_{j-1}/\mu_j$。

假定序列收敛，我们需要 $\sum_{i=0}^{\infty}P_i=1$。用 $r_j\triangleq\rho_j\cdots\rho_2\rho_1$ 和 $r_0=1$，这意味着 $P_0\sum_{i=0}^{\infty}r_i=1$ 或 $P_0=1/\sum_{i=0}^{\infty}r_i$。因此生–死马尔可夫链的稳态概率为

$$
P_j=r_j\Big/\sum_{i=0}^{\infty}r_i,\quad j\geqslant 0
$$

如果分母不收敛则意味着没有稳定状态，因此稳态的概率为0。这个模型一般称为 $M/M/1$ 队列。

2. 生–死参数为常数和有限存储长度 L 的 $M/M/1$ 队列 在这里，假定对所有的 i 有 $\lambda_i=\lambda$ 和 $\mu_i=\mu$，队列的长度不能超过 L。这个随机模型可以应用到有限长缓冲区分析，如图9.2-10所示。动态方程为

$$
\begin{aligned}
\mathrm{d}P_0(t)/\mathrm{d}t&=-\lambda P_0(t)+\mu P_1(t)\\
\mathrm{d}P_1(t)/\mathrm{d}t&=+\lambda P_0(t)-(\lambda+\mu)P_1(t)+\mu P_2(t)\\
&\vdots\quad\vdots\quad\vdots\\
\mathrm{d}P_L(t)/\mathrm{d}t&=+\lambda P_{L-1}(t)-\mu P_L(t)
\end{aligned}
$$

我们注意到，第一个和最后一个方程中只包含两项，因为死亡不能出现在一个空的队列中，同样当队列已经达到最大尺寸 L 时出生也不可能发生。从这些方程，很容易获得稳态解为：当 $0\leqslant i\leqslant L$ 时，$P_i=\rho^iP_0$，其中 $\rho\triangleq\lambda/\mu$。因为缓冲区必须处在某种状态，我们得到 $\sum_{i=0}^{L}\rho^iP_0=1$，或 $P_0=(1-\rho)/(1-\rho^{L+1})$。当缓冲区满时发生饱和。这一事件的稳态概率 $P_L=\rho^L(1-\rho)/(1-\rho^{L+1})$。因此，当出生率是死亡率的一半、以及缓冲区的大小为10时，发生饱和的概率大概为 5×10^{-4}。

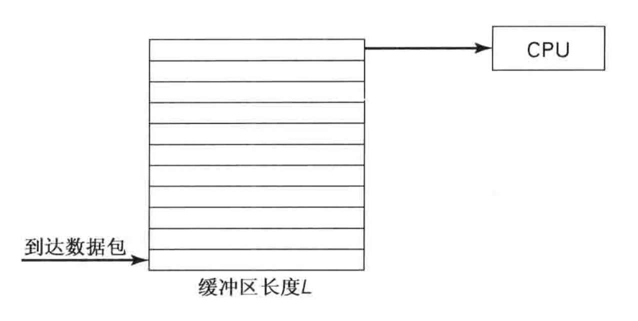


图 9.2-10 数据包到达具有有限长尺寸L的缓冲区的示意图

例 9.2-5 (平均队列大小)在计算机和通信网络中，分组交换是指数据包从节点到节点的块传输。在每个节点，数据包的处理主要是确定在源-目的地的路径中下一个链路。数据包的到达时间、它们必须在缓冲区中等待的时间，以及在 CPU(中央处理单元)中的服务时间都是随机变量。

假定先到先服务，无限容量的缓冲区，服务时间服从参数为 μ 的指数分布，在每个时间单元，到达数服从率参数为 λ 的泊松分布。我们从这一节的前面部分知道泊松过程的到达间隔时间是独立同分布的随机变量，服从参数为 λ 的指数分布。这种情况的状态图与图 9.2-9 相同，除了 $\mu_1=\mu_2=\cdots=\mu$ 和 $\lambda_0=\lambda_1=\cdots=\lambda$。将之前讨论的结果应用在这个例子上，可求得 $P_i=\rho^iP_0$，对所有 $0\leqslant i$ 成立，其中 $\rho\triangleq\lambda/\mu$，$P_0=(1-\rho)$。因此，在稳态时 $P_i=\rho^i(1-\rho)$，队列中数据包的平均数量 $E[N]$ 的计算结果为

$$E[N]=\sum_{i=0}^{\infty}iP_i=\frac{\rho}{(1-\rho)^2}$$

我们把详细的计算过程留给大家作为练习。

例 9.2-6 (有限容量缓冲区)重温例 9.2-5，区别是现在到达的数据包存储在一个大小为 L 的缓冲内。考虑下面的设置：存储在缓存中的数据是由基于先到先服务原则的 CPU 处理的。

假定在时刻 t 缓冲区是满的，此时在 CPU 中有一个数据包在处理，而且有一个数据包正在到达缓冲区。如果这个包和前一个包之间的间隔时间 τ_i 小于 τ_s，τ_s 数据包在 CPU 中的服务时间，则到达的数据包会丢失。这一事件发生的概率为

$$\begin{aligned}P[\text{“数据包丢失”}]&=P[\text{“饱和状态”}\cap\{\tau_s>\tau_i\}]\\&=\rho^L(1-\rho)/(1-\rho^{L+1})\times P[\tau_s-\tau_i>0]\end{aligned}$$

由于事件的“饱和状态”和$\{\tau_s>\tau_i\}$是独立的。由于 τ_s 和 τ_i 是独立的，则概率 $P[\tau_s-\tau_i>0]$ 可以通过卷积很容易计算出来。结果是 $P[\tau_s-\tau_i>0]=\lambda/(\lambda+\mu)$。那么，丢失即将到达的数据包的概率为

$$P[\text{“数据包丢失”}]=\rho^L(1-\rho)/(1-\rho^{L+1})\times\rho/(1+\rho)$$

如果 $\rho=0.5$，则对于尺寸为 10 的缓冲区，$P[\text{“数据包丢失”}]=1.6\times10^{-4}$，其中到达率等于服务率的一半。

查普曼-科尔莫戈罗夫方程

在第 8 章马尔可夫随机序列的例子中，我们定义的转移概率密度为一步转移，也就是从 $n-1$ 到 n。更一般地，我们可以定义从时间 n 到 $n+k$ 的转移概率密度作为一般马尔可夫随机序列的定义，其中 $k\geqslant0$。然而，在这个更一般的情况下，必须确保这个多步转移概率密度是一致的，也就是存在一个一步概率密度按顺序产生相同的结果。这个问题在随机过程的情况

则更为重要，因为连续时间的情况这是可以看成多步转移概率密度，即在任何两个 $t_2 \neq t_1$ 的时间之间总是存在一个时间。

例如，给定一个连续时间转移密度 $f_X(x_2|x_1;t_2,t_1)$，我们怎样根据下列方程求出无条件 pdf $f_X(x;t)$？

$$f_X(x_2;t_2)=\int_{-\infty}^{+\infty}f_X(x_2|x_1;t_2,t_1)f_X(x_1;t_1)\mathrm{d}x_1$$

式中对所有 $t_2>t_1$ 和所有 x_1 和 x_2 都成立。

查普曼-科尔莫戈罗夫方程为这些一般的转移密度的求解给出了必要的、充分的条件。查普曼-科尔莫戈罗夫方程还有离散值情况下的版本，其中涉及了多步转移的 PMF。

考虑三个时间 $t_3>t_2>t_1$，以及马尔可夫过程在这三个时间的的随机变量 $X(t_3)$，$X(t_2)$ 和 $X(t_1)$。我们希望计算当给定 $X(t_1)$ 时 $X(t_3)$ 的条件密度。首先，将联合 pdf 写成

$$f_X(x_3,x_1;t_3,t_1)=\int_{-\infty}^{+\infty}f_X(x_3|x_2,x_1;t_3,t_2,t_1)f_X(x_2,x_1;t_2,t_1)\mathrm{d}x_2$$

如果将方程两边都除以 $f(x_1;t_1)$，可得到

$$f_X(x_3|x_1)=\int_{-\infty}^{+\infty}f_X(x_3|x_2,x_1)f_X(x_2|x_1)\mathrm{d}x_2$$

其中省略了时间 t_i，以使式子看起来简单。应用马尔可夫特性，上式变为

$$f_X(x_3|x_1)=\int_{-\infty}^{+\infty}f_X(x_3|x_2)f_X(x_2|x_1)\mathrm{d}x_2 \tag{9.2-22}$$

这就是查普曼-科尔莫戈罗夫方程，用来求解马尔可夫过程的转移密度 $f_X(x_3|x_1)$。该方程适用于所有 $t_3>t_2>t_1$ 和所有 x_3 和 x_1 的值。能够证明在式(9.2-22)中表述的查普曼-科尔莫戈罗夫条件[9-5]，对于转移密度的存在性问题也是充分的。

由随机序列产生随机过程

当对马尔可夫随机序列的模拟次数取极限，我们就可以得到一个马尔可夫随机过程。例如，考虑由以下方程产生的随机序列

$$X[n]=\rho X[n-1]+W[n],\ -\infty<n<+\infty$$

参见例 8.4-6，为了确保系统稳定，设 $|\rho|<1.0$，则可得到 $X[n]$ 的相关函数为

$$R_{XX}[m]=\sigma_W^2\,\rho^{|m|}$$

其中 σ_W^2 是独立随机序列 $W[n]$ 的方差。用 $X(nT)$ 代替 $X[n]$，并且当满足 $nT\leqslant t\leqslant(n+1)T$ 时设 $X(t)=X[nT]$，得到

$$R_{XX}(t+\tau,t)=\sigma_W^2\ \rho^{|\tau/T|}=\sigma_W^2\ \exp(-\alpha|\tau|)$$

其中 $\alpha\triangleq\dfrac{1}{T}\ln\dfrac{1}{\rho}$ 或 $\rho=\exp(-\alpha T)$。这样如果我们对应 $k=1,2,3,\cdots$，在 $T_k\triangleq T_0/k$ 时刻产生了一组模拟结果，则对每组模拟结果 $\rho_k\triangleq\sqrt[k]{\exp(-\alpha T_0)}$，可以得到一组越来越密集的趋近于随机过程 $X(t)$ 的近似值，它是 WSS，相关函数为

$$R_{XX}(t+\tau,t)=\sigma_W^2\ \exp(-\alpha|\tau|)$$

9.3 具有随机信号输入的线性连续系统

本节介绍随机过程的变换。我们关注有记忆的线性变换，因为无记忆的情况可以通过第3章随机变量的变换方法来处理。首先回顾线性连续时间系统的定义。

定义 9.3-1 设 $x_1(t)$ 和 $x_2(t)$ 是两个确定的时间函数，a_1 和 a_2 是两个标量常数。线性系统用方程描述为 $y=L\{x\}$，则当满足以下条件时，系统为线性

$$L\{a_1x_1(t)+a_2x_2(t)\}=a_1L\{x_1(t)\}+a_2L\{x_2(t)\} \tag{9.3-1}$$

对于所有函数 x_1 和 x_2 以及所有标量 a_1 和 a_2 都成立。

这等于说，一个输入的加权和的响应必须等于每一个单独的输入的响应的加权和。同时，我们注意到在这个定义中输入必须在系统(算子)L 允许的输入空间中。当我们考虑将 L 推广到随机过程输入时，最自然的选择是输入 X 的样本函数，然后求相应的输出样本函数，从而定义一个新的随机过程 Y。正如原随机过程 X 是从样本空间到函数空间的映射，线性系统再将这个函数空间映射到一个新的函数空间，两个映射的级联定义了输出随机过程。图 9.3-1 描绘了这个过程。这一节我们的目标是求经过线性系统后输出随机过程的一阶矩和二阶矩，即经过线性系统变换后的均值和相关函数(协方差)的值。

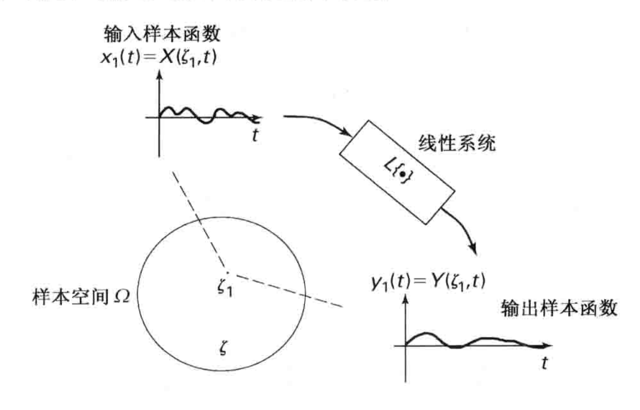


图 9.3-1 一个随机过程经过线性系统的图解

定理 9.3-1 设随机过程 $X(t)$ 经过线性系统 L 后产生输出过程 $Y(t)$。输出的均值为

$$\begin{aligned}E[Y(t)]&=L\{E[X(t)]\}\\&=L\{\mu_X(t)\}\end{aligned} \tag{9.3-2}$$

证明 通过定义可得到：对每个样本函数

$$Y(t,\zeta)=L\{X(t,\zeta)\}$$

于是

$$E[Y(t)]=E[L\{X(t)\}]$$

如果可以交换两个算子，则可以得到输出的均值函数是输入的均值函数经过线性变换 L 的结果。理由如下，如果假定线性算子 L 可以用叠加积分表示

$$Y(t)=\int_{-\infty}^{+\infty}h(t,\tau)X(\tau)\mathrm{d}\tau$$

取数学期望，可得到

$$E[Y(t)] = E\left[\int_{-\infty}^{+\infty} h(t,\tau)X(\tau)\mathrm{d}\tau\right]$$
$$= \int_{-\infty}^{+\infty} h(t,\tau)E[X(\tau)]\mathrm{d}\tau$$
$$= L\{\mu_X(t)\}$$

在第10章学习了均方随机积分后再对这个定理进行详细证明。现在假设它是有效的，然后看看线性系统输出相关函数的分析。现在有两个随机过程需要考虑，输入和输出，为此引入互相关函数为$E[X(t_1)Y^*(t_2)]$。我们定义互相关函数

$$R_{XY}(t_1,t_2) \triangleq E[X(t_1)Y^*(t_2)]$$

首先由输入的自相关函数$R_{XX}(t_1, t_2)$计算互相关函数$R_{XY}(t_1, t_2)$，然后计算输出的自相关函数$R_{YY}(t_1, t_2)$。如果输入过程的均值为零，则由定理9.3-1得到输出过程的均值也为零。因此，以下的结论对方差函数同样成立，只需将输入变为中心过程$X_c(t) \triangleq X(t) - \mu_X(t)$，就会产生中心输出$Y_c(t) \triangleq Y(t) - \mu_Y(t)$。

定理9.3-2　设$X(t)$和$Y(t)$为线性算子L的输入和输出随机过程，则

$$R_{XY}(t_1,t_2) = L_2^*\{R_{XX}(t_1,t_2)\} \tag{9.3-3}$$

$$R_{YY}(t_1,t_2) = L_1\{R_{XY}(t_1,t_2)\} \tag{9.3-4}$$

其中L_i是算子L对时间变量t_i做变换。

证明

$$X(t_1)Y^*(t_2) = X(t_1)L_2^*\{X^*(t_2)\}$$
$$= L_2^*\{X(t_1)X^*(t_2)\}$$

其中用了共轭算子L^*，其对应的冲激响应为$h^*(t, \tau)$，也就是$h(t, \tau)$的复共轭，则

$$E[X(t_1)Y^*(t_2)] = E\left[L_2^*\{X(t_1)X^*(t_2)\}\right]$$
$$= L_2^*\{E[X(t_1)X^*(t_2)]\} \quad \text{通过交换 } L_2^* \text{ 和 } E$$
$$= L_2^*\{R_{XX}(t_1,t_2)\}$$

这就是方程(9.3-3)。同样，为了证明方程(9.3-4)，乘以$Y^*(t_2)$，可得到

$$Y(t_1)Y^*(t_2) = L_1\{X(t_1)Y^*(t_2)\}$$

于是

$$E[Y(t_1)Y^*(t_2)] = E\left[L_1\{X(t_1)Y^*(t_2)\}\right]$$
$$= L_1\{E[X(t_1)Y^*(t_2)]\} \quad \text{通过交换 } L_1 \text{ 和 } E$$
$$= L_1\{R_{XY}(t_1,t_2)\}$$

这就是方程(9.3-4)。由方程(9.3-3)和方程(9.3-4)，可得到

$$R_{YY}(t_1,t_2) = L_1L_2^*\{R_{XX}(t_1,t_2)\} \tag{9.3-5}$$

例9.3-1　（边缘或“变化”检测器）设$X(t)$是一个实随机过程，可看成某传感器信号，定义$Y(t) \triangleq L\{X(t)\} \triangleq X(t) - X(t-1)$，则

$$E[Y(t)] = L\{\mu_X(t)\} = \mu_X(t) - \mu_X(t-1)$$

也就是

$$R_{XY}(t_1,t_2)=L_2\{R_{XX}(t_1,t_2)\}=R_{XX}(t_1,t_2)-R_{XX}(t_1,t_2-1)$$

和

$$\begin{aligned}R_{YY}(t_1,t_2)=L_1\{R_{XY}(t_1,t_2)\}&=R_{XY}(t_1,t_2)-R_{XY}(t_1-1,t_2)\\&=R_{XX}(t_1,t_2)-R_{XX}(t_1-1,t_2)-R_{XX}(t_1,t_2-1)\\&\quad+R_{XX}(t_1-1,t_2-1)\end{aligned}$$

如果 $\mu_X(t)=0$，并且

$$R_{XX}(t_1,t_2)\triangleq\sigma_X^2\exp(-\alpha|t_1-t_2|)$$

由于 $\mu_X=0$，则 $E[Y(t)]=0$，且

$$R_{YY}(t_1,t_2)=\sigma_X^2\left(2\exp(-\alpha|t_1-t_2|)-\exp(-\alpha|t_1-t_2-1|)-\exp(-\alpha|t_1-t_2+1|)\right)$$

我们注意到 R_{XX} 和 R_{XY} 都是两个观测时刻 t_1 和 t_2 差的函数。输入自相关函数 R_{XX} 如图9.3-2所示，其中 $\alpha=2$，$\sigma_X^2=2$。注意到图9.3-3中输出自相关函数的负值部分，是由边缘检测器的差分运算导致的。$Y(t)$ 的方差为常数，其表达式为

$$\sigma_Y^2(t)=\sigma_Y^2=2\sigma_X^2[1-\exp(-\alpha)]$$

我们看到随着 α 趋近于0，Y 的方差也趋近于0。这是因为随着 α 趋近于0，$X(t)$ 和 $X(t-1)$ 正相关性越来越强，因此它们的差的功率很小。

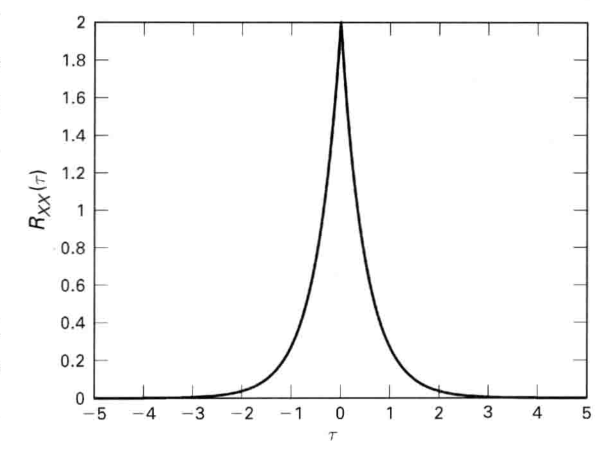


图9.3-2 例9.3-1的输入自相关函数 R_{XX}，其中 $\tau=t_1-t_2$

例9.3-2 (导数过程)设 $X(t)$ 为一个实过程，均值为常数 $\mu_X(t)=\mu$，协方差函数为

$$K_{XX}(t,s)=\sigma^2\cos\omega_0(t-s)$$

我们希望得到导数过程 $X'(t)$ 的均值和方差。这里的线性算子是 $\mathrm{d}(\cdot)/\mathrm{d}t$。首先计算均值

$$\mu_{X'}(t)=E[X'(t)]=\frac{\mathrm{d}}{\mathrm{d}t}E[X(t)]=\frac{\mathrm{d}}{\mathrm{d}t}\mu_X(t)=\frac{\mathrm{d}}{\mathrm{d}t}\mu=0$$

现在，对这个实随机过程，由于 $\mu'_X=0$，$X'(t)$ 的协方差函数为

$$K_{X'X'}(t_1,t_2)=E[X'(t_1)X'(t_2)]$$

通过式(9.3-5)和 $X'(t)=Y(t)$，得到

$$\begin{aligned}K_{X'X'}(t_1,t_2)&=\frac{\partial}{\partial t_1}\left(\frac{\partial}{\partial t_2}K_{XX}(t_1,t_2)\right)=\frac{\partial}{\partial t_1}\left(\frac{\partial}{\partial t_2}\sigma^2\cos\omega_0(t_1-t_2)\right)\\&=\frac{\partial}{\partial t_1}(\omega_0\sigma^2\sin\omega_0(t_1-t_2))=(\omega_0\sigma)^2\cos\omega_0(t_1-t_2)\end{aligned}$$

我们注意到，得到的结果刚好是原来的协方差函数乘了一个因子 ω_0^2。这种形式上的相似性是

由于$K_{XX}(t, s)$是具有随机振幅和相位的正弦波(参见例9.1-5)信号的协方差函数。由于相位是随机的，正弦及其导数余弦在形状上是难以区分的。

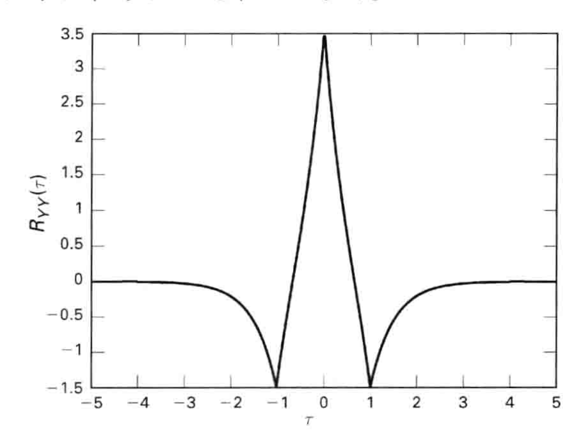


图9.3-3 例9.3-1的输出自相关函数R_{YY}与$\tau = t_1 - t_2$的图形

白噪声

考虑9.2节提到的维纳过程。在这里，我们考虑这一过程的导数。对任意$\alpha > 0$，维纳过程的协方差函数为$K_{XX}(t_1, t_2) = \alpha\min(t_1, t_2)$，均值$\mu_X = 0$。设$W(t) = \mathrm{d}X(t)/\mathrm{d}t$。继续分析上面的例子，可以计算得到$\mu_W(t) = E[\mathrm{d}X(t)/\mathrm{d}t] = \mathrm{d}\mu_X(t)/\mathrm{d}t = 0$。协方差为

$$
\begin{aligned}
K_{WW}(t_1,t_2) &= \frac{\partial}{\partial t_1}\left(\frac{\partial}{\partial t_2}K_{XX}(t_1,t_2)\right) \\
&= \frac{\partial}{\partial t_1}\left(\frac{\partial}{\partial t_2}\alpha\min(t_1,t_2)\right) \\
&= \frac{\partial}{\partial t_1}\left(\frac{\partial}{\partial t_2}\begin{cases}\alpha t_2, & t_2 < t_1 \\ \alpha t_1, & t_2 \geqslant t_1\end{cases}\right) \\
&= \frac{\partial}{\partial t_1}\begin{cases}\alpha, & t_2 < t_1 \\ 0, & t_2 > t_1\end{cases} \\
&= \frac{\partial}{\partial t_1}\begin{cases}0, & t_1 < t_2 \\ \alpha, & t_1 > t_2\end{cases} \\
&= \frac{\partial}{\partial t_1}\alpha u(t_1 - t_2) \\
&= \alpha\delta(t_1 - t_2)
\end{aligned}
$$

因此，白噪声的协方差函数为冲激函数。由于白噪声的均值始终为0，则自相关函数也为冲激函数，显然

$$
R_{WW}(t_1,t_2) = \sigma^2\delta(t_1 - t_2) = K_{WW}(t_1,t_2) \tag{9.3-6}
$$

其中σ^2代替了α，但我们也要注意到这个过程的功率$E[|W(t)|^2] = \sigma^2\delta(0) = \infty$，而不是$\sigma^2$。事实上，$\sigma^2$是白噪声过程的功率谱密度。

注意到这里的样本函数是不连续的，而白噪声过程是不可分的①。

① 可分离的概念(参见9.1节)是指由t轴上的可数点集来决定信号的特性(如时间轴)。也就是一旦知道了一个随机过程的可数点集的概率密度，则这个过程的完整概率密度就知道了(参见参考文献[9-6])。

9.4 一些关于随机过程的有用分类

这里我们看几类随机过程和过程对，这些分类方法同样适用于之前研究过的随机序列。

定义 9.4-1 设 X 和 Y 为随机过程，它们是：

(a)不相关的，如果 $R_{XY}(t_1, t_2)=\mu_X(t_1)\mu_Y^*(t_2)$，对所有 t_1 和 t_2。

(b)正交的，如果 $R_{XY}(t_1, t_2)=0$，对所有 t_1 和 t_2。

(c)独立的，如果对所有正整数 n，X 和 Y 的 n 阶分布函数可分解为

$$\begin{aligned}&F_{XY}(x_1,y_1,x_2,y_2,\cdots,x_n,y_n;t_1,\cdots,t_n)\\&= F_X(x_1,\cdots,x_n;t_1,\cdots,t_n)F_Y(y_1,\cdots,y_n;t_1,\cdots,t_n)\end{aligned}$$

对所有 x_i，y_i 以及所有 t_1，…，t_n 都成立。

我们注意到如果两个随机过程是不相关的且至少有一个随机过程的均值为 0，则它们是正交的。事实上，正交的定义只有在随机过程均值为 0 时才有用，在这种情况下正交与不相关等价。正交的概念是第 5 章为随机向量引入的。这个概念在第 11 章中用来估计随机过程和序列时是非常有用的。

随机过程可以是不相关、正交或独立的。例如可以是 $R_{XX}(t_1, t_2)=0$，对所有 $t_1\neq t_2$，在这种情况下我们称 $X(t)$ 为正交随机过程。同样 $X(t)$ 与 $\{X(t_1), \cdots, X(t_n)\}$ 是相互独立的，对所有 $t\notin\{t_1, \cdots, t_n\}$ 和所有 t_1，…，t_n 成立，其中 $n\geqslant1$，则我们称 $X(t)$ 为独立的随机过程。显然，由于时间上任意小的变化都引起完全的独立，所以这样的随机过程的样本函数变化会非常剧烈。

平稳性

当一个随机过程的统计特性不随时间改变，我们称其为平稳随机过程。正式的定义是：

定义 9.4-2 如果一个随机过程 $X(t)$ 的 n 阶分布函数与 $X(t+T)$ 相同，则称其为平稳随机过程。也就是两个 n 维函数

$$F_X(x_1,\cdots,x_n;t_1,\cdots,t_n)=F_X(x_1,\cdots,x_n;t_1+T,\cdots,t_n+T)$$

对于所有 T 以及所有正整数 n 和所有 t_1，…，t_n 都相同。

如果分布函数是可微的，则上式可以用概率密度等价为

$$f_X(x_1,\cdots,x_n;t_1,\cdots,t_n)=f_X(x_1,\cdots,x_n;t_1+T,\cdots,t_n+T)$$

这是运用更广泛地判断随机过程是否平稳的条件。此定义隐含了平稳随机过程的均值为常数。为了证明这个结论，我们注意到对所有 T，有 $f(x; t)=f(x; t+T)$ 成立，如果设 $T=-t$，则 $f(x; t)=f(x; 0)$，则可以得到 $E[X(t)]=\mu_X(t)=\mu_X(0)$ 为常数。

由于二阶概率密度位移不变性，即

$$f(x_1,x_2;t_1,t_2)=f(x_1,x_2;t_1+T,t_2+T)$$

同样令 $T=-t_2$，则

$$f(x_1,x_2;t_1,t_2)=f(x_1,x_2;t_1-t_2,0)$$

这意味着 $E[X(t_1)X^*(t_2)]=R_{XX}(t_1-t_2, 0)$。因此，在平稳条件下，相关函数只与两个采样

时间之间的时间差 $\tau \triangleq t_1 - t_2$ 有关。因此，我们定义一个参数的相关函数：

$$\begin{aligned} R_{XX}(\tau) &\triangleq R_{XX}(\tau, 0) \\ &= E[X(t+\tau)X^*(t)] \end{aligned} \tag{9.4-1}$$

其在函数关系上与参数 t 无关。这种相关函数的例子已在 9.3 节中见过。

如果不直接关注 n 阶分布函数，而仅仅考虑一阶和二阶矩，则我们得到了广义平稳的定义。

定义 9.4-3　一个随机过程 X 如果满足 $E[X(t)] = \mu_X$ 以及 $E[X(t+\tau)X^*(t)] = R_{XX}(\tau)$ 与时间参数 t 无关，其中 $-\infty < \tau < +\infty$，则称其为广义平稳随机过程(WSS)。

例 9.4-1　(广义平稳的复指数过程)设 $X(t) \triangleq A\exp(\mathrm{j}2\pi ft)$，其中 f 为一已知的实常数，A 为均值为 0 的平均功率 $E[A^2]$ 有限的实随机变量。计算 $X(t)$ 的均值和相关函数，可得到

$$E[X(t)] = E[A\exp(\mathrm{j}2\pi ft)] = E[A]\exp(\mathrm{j}2\pi ft) = 0$$

和

$$E[X(t+\tau)X^*(t)] = E[A\exp(\mathrm{j}2\pi f(t+\tau))A\exp(-\mathrm{j}2\pi ft)] = E[A^2]\exp(\mathrm{j}2\pi f\tau) = R_{XX}(\tau)$$

注意到这里 $E[A] = 0$ 是 WSS 的必要条件。问题：如果不是复指数函数而是一个余弦函数，均值和相关函数又是怎样的？

例 9.4-1 中的过程，虽然是广义平稳的，但实际上却不平稳。例如 $X(0)$ 为纯实数，而且 $X(1/(4f))$ 总是纯虚数。我们因此得到结论：广义平稳的特性要弱于平稳的特性。

我们把这个例子推广到 M 个复正弦信号，可以得到零均值广义平稳过程基本的频域表示。考虑

$$X(t) = \sum_{k=1}^{M} A_k \exp(\mathrm{j}2\pi f_k t)$$

其中复随机变量 A_k 是不相关的，均值为 0，方差为 σ_k^2。则 $X(t)$ 为 WSS，均值为 0，自相关或自协方差函数等于

$$R_{XX}(\tau) = \sum_{k=1}^{M} \sigma_k^2 \exp(\mathrm{j}2\pi f_k \tau) \tag{9.4-2}$$

对这个随机过程 $X(t)$，随机系数 $\{A_k\}$ 构成了其频域表示。由确定函数的傅里叶分析可以推断出随着 M 增大以及 f_k 变密，也就是，f_k 之间的空间变小，它覆盖了我们感兴趣的频率范围，大多数随机过程会有这样一个近似的表示。这种情况参见 10.6 节。

9.5　广义平稳随机过程和 LSI 系统

在这一节我们分析联合平稳且是二阶的随机过程，即

$$E[|X(t)|^2] < \infty$$

平稳二阶过程的自相关和互相关函数的某些重要特性总结如下。当然，它们对各自的相关函数同样成立。

(1) $|R_{XX}(\tau)| \leqslant R_{XX}(0)$，对于实数的情况，可直接由 $E[|X(t+\tau)-X(t)|^2] \geqslant 0$ 得到。

(2) $|R_{XX}(\tau)| \leqslant \sqrt{R_{XX}(0)R_{YY}(0)}$，这可由施瓦茨不等式得到(参见 4.3 节)。它也可以证明(1)中的复数情况。

(3) $R_{XX}(\tau)=R_{XX}^*(-\tau)$，这是由于对 WSS，有 $E[X(t+\tau)X^*(t)]=E[X(t)X^*(t-\tau)]=E[X(t-\tau)X^*(t)]$，这个特性被称为共轭对称性。对于实过程，则上述特性变成了偶对称，即 $R_{XX}(\tau)=R_{XX}(-\tau)$。对于复值平稳随机过程，其自相关函数的另一个重要特性是它必须是半正定的(即下一条特性)。

(4) 对于所有 $N>0$ 和 $t_1<t_2<\cdots<t_N$ 以及所有的复数 $a_1, a_2, \cdots, a_N$，满足

$$\sum_{k=1}^{N}\sum_{l=1}^{N} a_k a_l^* R_{XX}(t_k-t_l) \geqslant 0$$

这是在 9.1 节中证明过的给定一个函数 $g(t,s)=g(t-s)$ 是相关函数的必要条件。我们还将证明它同样是一个充分条件，因此半正定实际上是自相关函数的特征。但是通常直接证明特性(4)会非常困难。

从定理 9.3-1 和定理 9.3-2 出发，可以推导 LSI 系统的一般情况。重写式(9.3-2)，可得到

$$\begin{aligned} E[Y(t)] &= L\{\mu_X(t)\} \\ &= \int_{-\infty}^{\infty} \mu_X(\tau)h(t-\tau)\,\mathrm{d}\tau \\ &= \mu_X(t) * h(t) \end{aligned}$$

运用定理 9.3-2 和式(9.3-3)、式(9.3-4)，有

$$R_{XY}(t_1,t_2)=\int_{-\infty}^{+\infty} h^*(\tau_2)R_{XX}(t_1,t_2-\tau_2)\mathrm{d}\tau_2$$

和

$$R_{YY}(t_1,t_2)=\int_{-\infty}^{+\infty} h(\tau_1)R_{XY}(t_1-\tau_1,t_2)\mathrm{d}\tau_1$$

可以用卷积算子表示为

$$R_{XY}(t_1,t_2)=h^*(t_2) * R_{XX}(t_1,t_2)$$

其中卷积算子沿 t_2 轴，而

$$R_{YY}(t_1,t_2)=h(t_1) * R_{XY}(t_1,t_2)$$

其中卷积是沿 t_1 轴将这两个方程组合起来，我们得到 $R_{YY}(t_1, t_2)=h(t_1) * R_{XX}(t_1, t_2) * h^*(t_2)$。

广义平稳情况

如果将平稳随机过程 $X(t)$ 输入到一个冲激响应为 $h(t)$ 的 LSI 系统，则输出随机过程可以用卷积积分表示为

$$Y(t)=\int_{-\infty}^{+\infty} h(\tau)X(t-\tau)\mathrm{d}\tau \tag{9.5-1}$$

当然要满足积分存在的条件。计算输出过程 $Y(t)$ 的均值，可得到

$$
\begin{aligned}
E[Y(t)] &= \int_{-\infty}^{+\infty} h(\tau)E[X(t-\tau)]\mathrm{d}\tau \qquad \text{由定理9.3-1} \\
&= \int_{-\infty}^{+\infty} h(\tau)\mu_X \mathrm{d}\tau = \mu_X \int_{-\infty}^{+\infty} h(\tau)\mathrm{d}\tau \\
&= \mu_X H(0)
\end{aligned} \tag{9.5-2}
$$

其中 $H(\omega)$ 是系统的频率响应。

因此，我们看到输出的均值为常数并且等于输入的均值乘以系统函数在 $\omega=0$ 处的取值，$H(0)$称为系统的直流增益。如果计算输入和输出随机过程的互相关函数，可得到

$$
\begin{aligned}
R_{YX}(\tau) &= E[Y(t+\tau)X^*(t)] \\
&= E[Y(t)X^*(t-\tau)] \quad \text{用 } t-\tau \text{ 代替 } t \\
&= \int_{-\infty}^{+\infty} h(\alpha)E[X(t-\alpha)X^*(t-\tau)]\mathrm{d}\alpha
\end{aligned}
$$

运用定理9.3-2，将均值算子 E 放到积分式中，则

$$
= \int_{-\infty}^{+\infty} h(\alpha)R_{XX}(\tau-\alpha)\mathrm{d}\alpha
$$

其可以被重写为

$$
R_{YX}(\tau) = h(\tau) * R_{XX}(\tau) \tag{9.5-3}
$$

因此，互相关函数 R_{YX}等于 h 与自相关函数 R_{XX}的卷积。这个性质可以用来辨识未知系统（参见习题9.28）。

输出自相关函数 $R_{YY}(\tau)$可由互相关函数 $R_{YX}(\tau)$得到

$$
\begin{aligned}
R_{YY}(\tau) &= E[Y(t+\tau)Y^*(t)] \\
&= E[Y(t)Y^*(t-\tau)] \quad \text{用 } t \text{ 代替 } t-\tau \\
&= \int_{-\infty}^{+\infty} h^*(\alpha)E[Y(t)X^*(t-\tau-\alpha)]\mathrm{d}\alpha \\
&= \int_{-\infty}^{+\infty} h^*(\alpha)E[Y(t)X^*(t-(\tau+\alpha))]\mathrm{d}\alpha \\
&= \int_{-\infty}^{+\infty} h^*(\alpha)R_{YX}(\tau+\alpha)\mathrm{d}\alpha \\
&= \int_{-\infty}^{+\infty} h^*(-\alpha)R_{YX}(\tau-\alpha)\mathrm{d}\alpha \\
&= h^*(-\tau) * R_{YX}(\tau)
\end{aligned}
$$

组合两个方程，得到

$$
R_{YY}(\tau) = h(\tau) * h^*(-\tau) * R_{XX}(\tau) \tag{9.5-4}
$$

我们观察到当 $R_{XX}(\tau)=\delta(\tau)$时，输出的相关函数为$R_{YY}(\tau)=h(\tau)*h^*(-\tau)$，其有时被称为自相关冲激响应（AIR），表示为 $g(\tau)=h(\tau)*h^*(-\tau)$。注意 $g(\tau)$必须是半正定的，并且 $\mathrm{FT}\{g(\tau)\}=|H(\omega)|^2\geqslant 0$。

同样，我们也求得（证明留给读者作为练习）

$$R_{XY}(\tau)=\int_{-\infty}^{+\infty}h^*(-\alpha)R_{XX}(\tau-\alpha)\mathrm{d}\alpha$$
$$=h^*(-\tau)*R_{XX}(\tau) \tag{9.5-5a}$$

和

$$R_{YY}(\tau)=\int_{-\infty}^{+\infty}h(\alpha)R_{XY}(\tau-\alpha)\mathrm{d}\alpha$$
$$=h(\tau)*R_{XY}(\tau)$$
$$=h(\tau)*h^*(-\tau)*R_{XX}(\tau)$$
$$=g(\tau)*R_{XX}(\tau)$$

这个简练的式子是下列式子的简写：

$$R_{YY}(\tau)=\int_{-\infty}^{\infty}g(\tau')R_{XX}(\tau-\tau')\mathrm{d}\tau' \quad (卷积) \tag{9.5-5b}$$

$$g(\tau')=\int_{-\infty}^{\infty}h^*(\alpha)h(\alpha+\tau')\mathrm{d}\alpha \quad (相关积) \tag{9.5-5c}$$

例 9.5-1 (WSS 过程的导数)设二阶随机过程 $X(t)$ 是平稳的，相关函数为 $R_x(\tau)$，均值 $\mu_X(t)=\mu_X$ 为常数。系统为一导数运算，即

$$Y(t)=\frac{\mathrm{d}X(t)}{\mathrm{d}t}$$

运用上面的方程，我们得到 $\mu_Y(t)=\mathrm{d}\mu_X(t)/\mathrm{d}t=0$，互相关函数为

$$R_{XY}(\tau)=u_1^*(-\tau)*R_{XX}(\tau)$$
$$=-\frac{\mathrm{d}R_{XX}(\tau)}{\mathrm{d}\tau}$$

由于求导运算的冲激响应为 $h(t)=\mathrm{d}\delta(t)/\mathrm{d}t=u_1(t)$，冲激函数的导数 $\delta(t)$ 有时称为单位冲激偶①。

$$R_{YY}(\tau)=u_1(\tau)*R_{XY}(\tau)$$
$$=\frac{\mathrm{d}R_{XY}(\tau)}{\mathrm{d}\tau}$$
$$=-\frac{\mathrm{d}^2R_{XX}(\tau)}{\mathrm{d}\tau^2}$$

注意到这里 AIP 函数是 $g(\tau)=-u_2(\tau)$，是 $\delta(\tau)$ 的二阶导数的负值。

功率谱密度

对于广义平稳过程，我们定义其平均功率在频率上的密度，称为功率谱密度(psd)。

定义 9.5-1 设 $R_{XX}(\tau)$ 为自相关函数。我们定义功率谱密度为其傅里叶变换(如果它存在)，即

$$S_{XX}(\omega)\triangleq\int_{-\infty}^{+\infty}R_{XX}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau \tag{9.5-6}$$

① 在这个 u 函数标记中，$u_{-1}(t)=u(t)$ 为单位阶跃函数，$u_0(t)=\delta(t)$ 为单位冲激函数[9-9]。

在相当一般的条件下，它的逆变换等于 $R_{XX}(\tau)$（所有连续点）

$$R_{XX}(\tau)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}S_{XX}(\omega)\mathrm{e}^{+\mathrm{j}\omega\tau}\,\mathrm{d}\omega \tag{9.5-7}$$

我们标记

$$S_{XX}=\mathrm{FT}\{R_{XX}\}$$

和

$$R_{XX}=\mathrm{IFT}\{S_{XX}\}$$

其中 FT 和 IFT 分别代表傅里叶变换和傅里叶逆变换。

功率谱这个词稍后再做解释。我们目前将这定义为 $R_{XX}(\tau)$ 的傅里叶变换。也可以定义互相关函数 $R_{XY}(\tau)$ 的傅里叶变换从而得到一个频率的函数，我们称其为互功率谱密度

$$S_{XY}(\omega)\triangleq\int_{-\infty}^{+\infty}R_{XY}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}\tau \tag{9.5-8}$$

以后会看到功率谱 $S_{XX}(\omega)$ 是实的且处处为非负数，而且正如其名称所指，是随机信号的平均功率在频率上分布的密度函数。而互谱密度则没有这样的含义，而且通常为复数。

下面列举了一些功率谱密度 $S_{XX}(\omega)$ 的特性：

1. 由于 $R_{XX}(\tau)$ 共轭对称，所以 $S_{XX}(\omega)$ 是实数。
2. 如果 $X(t)$ 是实 WSS 过程，则由于 $R_{XX}(\tau)$ 是实的偶函数，所以 $S_{XX}(\omega)$ 也是偶函数。否则 $S_{XX}(\omega)$ 可能不是 ω 的偶函数。
3. $S_{XX}(\omega)\geqslant 0$（在定理 9.5-1 中说明）。

有关功率谱密度的其他性质在表 9.5-1 中说明。大家可以继续扩展这张表，但到此我们的目的已经达到了。注意这个运算在频域上的简单性。这提示我们对 LSI 系统和平稳或 WSS 过程，可以首先将输入自相关函数转换到频域上，进行指示运算（indicated operation），然后再将其逆变换到相关域来得到输出的相关函数。这与确定线性系统理论中的平移-不变系统的情况完全类似。

表 9.5-1 相关函数的特性及其对应的功率谱密度

随机过程	相关函数	功率谱密度
$X(t)$	$R_{XX}(\tau)$	$S_{XX}(\omega)$
$aX(t)$	$\|a\|^2R_{XX}(\tau)$	$\|a\|^2S_{XX}(\omega)$
$X_1(t)+X_2(t)$（X_1 和 X_2 正交）	$R_{X_1X_1}(\tau)+R_{X_2X_2}(\tau)$	$S_{X_1X_1}(\omega)+S_{X_2X_2}(\omega)$
$X'(t)$	$-\mathrm{d}^2R_{XX}(\tau)/\mathrm{d}\tau^2$	$\omega^2S_{XX}(\omega)$
$X^{(n)}(t)$	$(-1)^n\mathrm{d}^{2n}R_{XX}(\tau)/\mathrm{d}\tau^{2n}$	$\omega^{2n}S_{XX}(\omega)$
$X(t)\exp(\mathrm{j}\omega_0t)$	$\exp(\mathrm{j}\omega_0\tau)R_{XX}(\tau)$	$S_{XX}(\omega-\omega_0)$
$X(t)\cos(\omega_0t+\Theta)$ 其中 Θ 与 $X(t)$ 独立，且在 $[-\pi,+\pi]$ 上均匀分布	$\frac{1}{2}R_{XX}(\tau)\cos(\omega_0\tau)$	$\frac{1}{4}[S_{XX}(\omega+\omega_0)+S_{XX}(\omega-\omega_0)]$
$X(t)+b\quad(E[X(t)]=0)$	$R_{XX}(\tau)+\|b\|^2$	$S_{XX}(\omega)+2\pi\|b\|^2\delta(\omega)$

另一种情况是如果将 $S_{XX}(\omega)$ 描述为平均功率密度的解释是正确的话，则常数或均值分量的所有平均功率都集中在 $\omega=0$ 处，正如表中最后一行所示。而正如表倒数第 2 行所示，频率 ω_0 的调制意味着平均功率的分布在频域上平移了 ω_0，这两个结论应该是非常直观的。

例9.5-2 (白噪声的功率谱密度)参数为 σ^2 的白噪声 $W(t)$ 的相关函数为 $R_{WW}(\tau)=\sigma^2\delta(\tau)$，它的傅里叶变换，即功率谱密度为

$$S_{WW}(\omega)=\sigma^2,\quad -\infty<\omega<+\infty$$

功率谱密度是平坦的，因此“白噪声”名称的由来源于白光，因为白光在所有的波长上的功率相等。与白光类似，白噪声是理想化的，由于之前我们已经得到 $P_{WW}(0)=\infty$，意味着功率无穷大，所以其在物理世界里并不真实存在。此外，应该注意在白噪声情况下，参数 σ^2 应被理解为功率密度。

功率谱密度的解释

给定一个平稳随机过程 $X(t)$，考虑有限支撑段

$$X_T(t)\triangleq X(t)I_{[-T,+T]}(t)$$

其中 $I_{[-T,+T]}$ 是一个指示函数，如果 $-T\leqslant t\leqslant +T(T>0)$，则其值为1，否则为0。计算 X_T 的傅里叶变换得到

$$\mathrm{FT}\{X_T(t)\}=\int_{-T}^{+T}X(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t$$

这个随机变量的幅度平方值为

$$|\mathrm{FT}\{X_T(t)\}|^2=\int_{-T}^{+T}\int_{-T}^{+T}X(t_1)X^*(t_2)\mathrm{e}^{-\mathrm{j}\omega(t_1-t_2)}\mathrm{d}t_1\,\mathrm{d}t_2$$

除以 $2T$，取均值后得到

$$\frac{1}{2T}E\left[|\mathrm{FT}\{X_T(t)\}|^2\right]=\frac{1}{2T}\int_{-T}^{+T}\int_{-T}^{+T}R_{XX}(t_1-t_2)\mathrm{e}^{-\mathrm{j}\omega(t_1-t_2)}\,\mathrm{d}t_1\,\mathrm{d}t_2 \tag{9.5-9a}$$

为了计算等式右边的二重积分，引入新的坐标系统 $s=t_1+t_2$，$\tau=t_1-t_2$，(s,τ) 和 (t_1,t_2) 坐标的关系显示在图9.5-1(a)中。这个变换的雅可比值是1/2，积分区域是一个菱形 $\wp$，如图9.5-1(b)所示，它是将图9.5-1(a)顺时针旋转45°得到的，其边长为 $T\sqrt{2}$，于是式(9.5-9a)中的双重积分就变为

$$\begin{aligned}
&\frac{1}{4T}\iint_{\wp}R_{XX}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}\tau\,\mathrm{d}s\\
&=\frac{1}{4T}\left\{\int_{-2T}^{0}R_{XX}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\left[\int_{-(2T+\tau)}^{2T+\tau}\mathrm{d}s\right]\mathrm{d}\tau\right\}\\
&\quad+\frac{1}{4T}\left\{\int_{0}^{2T}R_{XX}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\left[\int_{-(2T-\tau)}^{2T-\tau}\mathrm{d}s\right]\mathrm{d}\tau\right\}=\int_{-2T}^{+2T}\left[1-\frac{|\tau|}{2T}\right]R_{XX}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau
\end{aligned}$$

如果 $T\to+\infty$，则这个积分趋向于式(9.5-6)，即

$$S_{XX}(\omega)=\lim_{T\to\infty}\frac{1}{2T}E\left[|\mathrm{FT}\{X_T(t)\}|^2\right] \tag{9.5-9b}$$

所以，$S_{XX}(\omega)$ 为实的非负函数，并且与频率 ω 处的平均功率有关。

接下来我们看两个用相关函数来计算功率谱密度的例子。

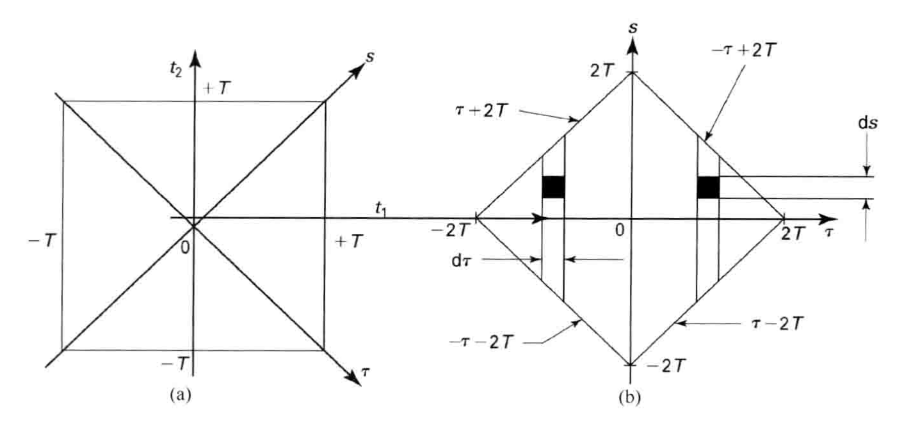


图 9.5-1 (a)在(t_1, t_2)平面内的方形区域；(b)经过$s=t_1+t_2$, $\tau=t_1-t_2$变换后菱形范围内的积分

例 9.5-3 求下面以指数形式表示的相关函数的功率谱密度，其中参数$\alpha>0$

$$R_{XX}(\tau)=\exp(-\alpha|\tau|),\quad -\infty<\tau<+\infty$$

这是 9.2 节讨论过的随机电报信号(RTS)的自相关函数，它的功率谱密度为

$$\begin{aligned}S_{XX}(\omega)&=\int_{-\infty}^{+\infty}R_{XX}(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau=\int_{-\infty}^{+\infty}\mathrm{e}^{-\alpha|\tau|}\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau\\&=\int_{-\infty}^{0}\mathrm{e}^{(\alpha-\mathrm{j}\omega)\tau}\mathrm{d}\tau+\int_{0}^{\infty}\mathrm{e}^{-(\alpha+\mathrm{j}\omega)\tau}\mathrm{d}\tau\\&=2\alpha/[\alpha^2+\omega^2],\quad -\infty<\omega<+\infty\end{aligned}$$

图 9.5-2 画出了当$\alpha=3$时的函数图。我们看到峰值出现在原点，等于$2/\alpha$。这个过程的 3 dB 带宽为α(如果S_{XX}确定是功率密度的话，将会证明这一点)。我们注意到相关函数R_{XX}在原点有一个尖头，而功率谱密度S_{XX}无尖头。事实上，S_{XX}是连续的并且处处可微的(如果R_{XX}是绝对可积的，则S_{XX}总是连续的)。

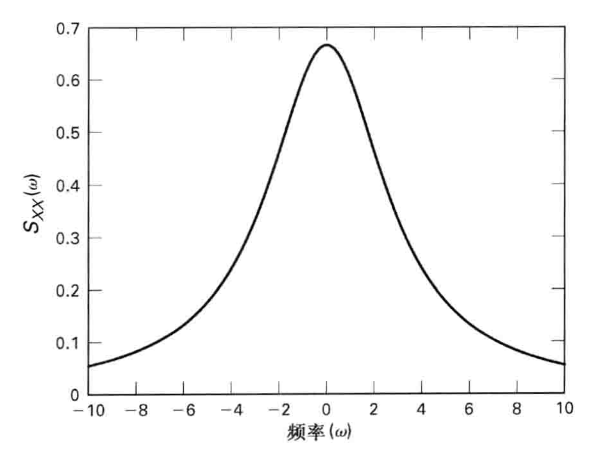


图 9.5-2 指数型相关函数的功率谱密度

图 9.5-2 是由以下的 MATLAB 程序画出来的图。

```
clear alpha=3;
b = [1.0 0.0 alpha^2];
w = linspace(-10,+10);
```

```
den = polyval(b,w);
num = 2*alpha;
S = num./den;
plot (w,S)
```

我们注意到功率谱密度衰减得非常慢，因此随机电报信号需要一个相对比较大的通带。功率谱密度尾部较长的原因是 RTS 的样本函数的跳变。

例 9.5-4 (三角形自相关函数的功率谱密度)考虑自相关函数为三角形形式，则相关函数在偏移 T 后为 0，其中 $T>0$

$$R_{XX}(\tau)=\max\left[1-\frac{|\tau|}{T},0\right]$$

在 9.2 节介绍的异步二进制信号(ABS)就出现了这样的相关函数。图 9.5-3 画出了这个函数。如果我们认识到这个三角函数可以看成宽度为 T，高度为 $1/\sqrt{T}$ 的两个矩形脉冲的卷积，则可以运用傅里叶变换的卷积定理[9-3, -4]来得到这个三角形自相关函数的功率谱密度，它刚好等于矩形脉冲的傅里叶变换(即 sinc 函数)的平方，矩形脉冲的变换为

$$\sqrt{T}\frac{\sin(\omega T/2)}{(\omega T/2)}$$

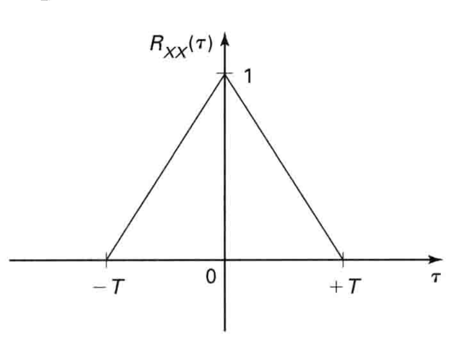


图 9.5-3 三角形自相关函数

这个三角形自相关函数的功率谱密度 S_{XX} 则为

$$S_{XX}(\omega)=T\left(\frac{\sin(\omega T/2)}{\omega T/2}\right)^2 \tag{9.5-10}$$

为了验证结果的正确性，我们注意到 $S_{XX}(0)$ 刚好是自相关函数的面积，对三角形很容易计算得到其结果为 T，因此

$$S_{XX}(0)=\int_{-\infty}^{+\infty}R_{XX}(\tau)\,\mathrm{d}\tau=2\cdot\frac{1}{2}\cdot 1\cdot T$$

出现三角形自相关函数的另一种情况是计算白噪声的积分平均值，即

$$X(t)\triangleq\frac{1}{\sqrt{T}}\int_{t-T}^{t}W(\tau)\,\mathrm{d}\tau$$

白噪声 $W(t)$ 的均值为 0、自相关函数为 $R_{WW}(\tau)=\delta(\tau)$。则 $\mu_X(t)=0$, $E[X(t_1)X(t_2)]$ 为

$$\begin{aligned}R_{XX}(t_1,t_2)&=\frac{1}{T}\int_{t_1-T}^{t_1}\int_{t_2-T}^{t_2}R_{WW}(s_1-s_2)\mathrm{d}s_1\,\mathrm{d}s_2\\&=\frac{1}{T}\int_{t_1-T}^{t_1}\left[\int_{t_2-T}^{t_2}\delta(s_2-s_1)\,\mathrm{d}s_2\right]\mathrm{d}s_1\end{aligned}$$

将方括号里面的积分定义为

$$g_{t_2}(s_1)\triangleq\int_{t_2-T}^{t_2}\delta(s_2-s_1)\mathrm{d}s_2=\begin{cases}1, & t_2-T\leqslant s_1\leqslant t_2\\0, & \text{其他}\end{cases}$$

为 s_1 的函数，图 9.5-4 显示了其图形，于是

$$R_{XX}(t_1,t_2)=\frac{1}{T}\int_{t_1-T}^{t_1}g_{t_2}(s_1)\mathrm{d}s_1$$
$$=\max\left[1-\frac{|t_1-t_2|}{T},0\right]$$

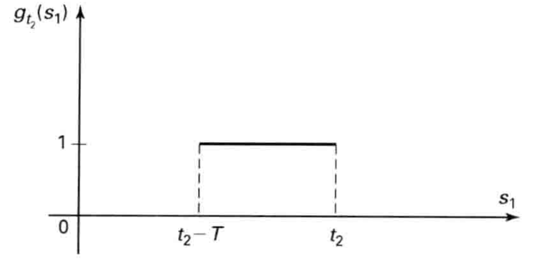


图9.5-4 当 $t_2>T$ 时方程相对于 s_1 的变化图

更多有关白噪声的讨论

白噪声的相关函数为冲激函数[参见式(9.3-6)]，则其功率谱密度为常数

$$S_{WW}(\omega)=\sigma^2,\qquad -\infty<\omega<+\infty$$

白噪声名称的由来是因为其功率谱在整个频率轴上为常数，与白光类似，白光的光谱对所有的波长都是恒定值①。这里我们把白噪声信号看成是二阶序列的极限。考虑一个均值为0的独立增量过程(参见定义9.2-1)，比如维纳过程 $[R_{XX}(t_1,t_2)=\sigma^2\min(t_1,t_2)]$ 或中心泊松过程 $N_C(t)=N(t)-\lambda t$，其相关函数为 $R_{N_CN_C}(t_1,t_2)=\lambda\min(t_1,t_2)$。事实上，这里只需要不相关的增量；即我们只需要 $X(t)$ 具有不相关的增量。对这类过程，根据式(9.2-17)，我们有

$$E\left[(X(t+\Delta)-X(t))^2\right]=\alpha\Delta$$

其中 α 为方差参数。

设 $X_\Delta(t)$ 表示一阶差分除以 Δ，即

$$X_\Delta(t)\triangleq[X(t+\Delta)-X(t)]/\Delta$$

我们有

$$E[X_\Delta^2(t)]=\alpha/\Delta$$

和

$$E[X_\Delta(t_1)X_\Delta(t_2)]=0,\qquad |t_2-t_1|>\Delta$$

如果考虑到 $|t_2-t_1|<\Delta$，则可以进行以下的计算，得出的相关函数刚好是例9.5-4中所示的三角函数。由于 $X(t_1+\Delta)-X(t_1)$ 服从 $N(0,\Delta)$ 分布，设 $t_1<t_2$，将 t_1 移动至0，t_2 移动至 t_2-t_1，则数学期望变成了

$$\frac{1}{\Delta^2}E\left[X(\Delta)\left(X(t_2-t_1+\Delta)-X(t_2-t_1)\right)\right]$$
$$=\frac{1}{\Delta^2}E\left[X(\Delta)\left(X(\Delta)-X(t_2-t_1)\right)\right]\qquad \text{由于 } (\Delta,t_2-t_1+\Delta]\cap(0,\Delta]=\phi$$
$$=\frac{1}{\Delta^2}[\alpha\Delta-\alpha(t_2-t_1)]=\frac{\alpha}{\Delta}[1-(t_2-t_1)/\Delta]$$

这样，这个由一阶差分生成的过程为平稳随机过程(均值为0)，相关函数 $R_{\Delta\Delta}(\tau)$ 为

$$R_{\Delta\Delta}(\tau)=\frac{\alpha}{\Delta}\max\left[1-\frac{|\tau|}{\Delta},0\right]$$

由图9.5-5我们注意到随着 Δ 趋近于0，这个相关函数趋近于一个冲激函数。

① 这是数学上的理想化模型！物理原理告诉我们，对于真实的模型，功率谱密度在 $\omega\to\infty$ 时必须趋近于0。

在例 9.5-4 中我们已经计算了一个三角函数的傅里叶变换，通过观察，可以将其功率谱密度写成

$$S_{\Delta\Delta}(\omega)=\alpha\left(\frac{\sin(\omega\Delta/2)}{\omega\Delta/2}\right)^2$$

这个功率谱在 $|\omega|=\pi/(3\Delta)$ 内近似平坦。随着 $\Delta\to 0$，$S_{\Delta\Delta}(\omega)$ 在所有频率点上都趋近于一个常数 α。因此随着 $\Delta\to 0$，$X_\Delta(t)$ “趋向于”白噪声，即不相关增量过程导数

$$\begin{aligned}R_{\dot{X}\dot{X}}(t_1,t_2)&=\frac{\partial^2}{\partial t_1\partial t_2}[\sigma^2\min(t_1,t_2)]\\&=\frac{\partial}{\partial t_1}[\sigma^2u(t_1-t_2)]\\&=\sigma^2\delta(t_1-t_2)\end{aligned}$$

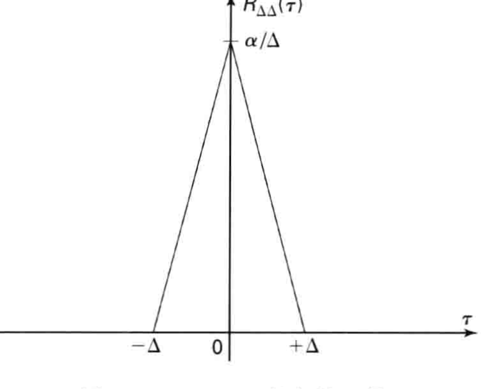


图 9.5-5　$X_\Delta(t)$ 的相关函数

如果一个系统对激励的响应是连续的，我们就说这个系统是连续的，即系统的运算是连续的运算。这意味着，输入有微小变化时输出也会产生微小变化。一个稳定的微分或差分方程就是这种连续运算的一个例子。我们看到对于用系统函数描述的线性时不变系统，当随机过程 $X_\Delta(t)$ 作用于这个系统，如果 Δ 很小，并且系统是低通系统，即系统函数随着 $|\omega|\to\infty$ 而趋向于0，则其响应只会有轻微的变化。因此用白噪声来构造这个极限输出就是最方便的方法(参见习题 9.36)。

如果将式(9.5-3)的两边分别做傅里叶变换，可得到互谱密度

$$S_{YX}(\omega)=H(\omega)S_{XX}(\omega) \tag{9.5-11}$$

由于 S_{YX} 是互相关表示 R_{YX} 的频域函数，式(9.5-11)告诉我们在 $H(\omega)$ 和 $S_{XX}(\omega)$ 的乘积较大时对应的频率处 $Y(t)$ 和 $X(t)$ 的互相关很强。类似地，由式(9.5-5)，可以得到

$$S_{XY}(\omega)=H^*(\omega)S_{XX}(\omega) \tag{9.5-12}$$

由式(9.5-4)，在这里重复一遍，即

$$R_{YY}(\tau)=h(\tau)*R_{XX}(\tau)*h^*(-\tau) \tag{9.5-13}$$

根据傅里叶变换，可得到

$$S_{YY}(\omega)=|H(\omega)|^2S_{XX}(\omega)=G(\omega)S_{XX}(\omega) \tag{9.5-14}$$

这两个公式是平稳随机过程理论中最重要的公式。特别地，式(9.5-14)说明了在某个频率处输出随机过程的平均功率是如何由输入的平均功率与 $|H(\omega)|^2$ 相乘得到的，我们称 $G(\omega)=|H(\omega)|^2$ 为功率谱传递函数。

例 9.5-5　(平均功率)一个 LSI 系统的传递函数为

$$H(\omega)=\operatorname{sgn}(\omega)\left(\frac{\omega}{2\pi}\right)^2\exp\left[-\mathrm{j}\left(\omega\cdot\frac{8}{\pi}\right)\right]W(\omega)$$

其中 sgn(·)是符号函数，频率窗函数 $W(\omega)$ 为

$$W(\omega)\triangleq\begin{cases}1, & |\omega|\leqslant 40\pi\\0, & \text{其他}\end{cases}$$

设输入平稳随机过程的自相关函数为

$$R_{XX}(\tau)=\frac{5}{2}\delta(\tau)+2$$

计算在0～1 Hz(单边带)频带内的平均可测功率，以弧度表示，则为双边带 $-2\pi\sim2\pi$ 范围内。首先对 $R_{XX}(\tau)$ 做傅里叶变换，得到 $S_{XX}(\omega)=\dfrac{5}{2}+4\pi\delta(\omega)$。然后计算功率传递函数 $G(\omega)=|H(\omega)|^2=\left(\dfrac{\omega}{2\pi}\right)^4W(\omega)$。则输出的功率谱密度为

$$S_{YY}(\omega)=\frac{5}{2}\left(\frac{\omega}{2\pi}\right)^4W(\omega)$$

输出的总平均功率为

$$R_{YY}(0)=\frac{1}{2\pi}\int_{-40\pi}^{+40\pi}\frac{5}{2}\left(\frac{\omega}{2\pi}\right)^4\mathrm{d}\omega$$

而在 $[-2\pi,2\pi]$ 频带内的功率为

$$P=\frac{1}{2\pi}\int_{-2\pi}^{+2\pi}\frac{5}{2}\left(\frac{\omega}{2\pi}\right)^4\mathrm{d}\omega$$
$$=1\ \mathrm{W}$$

从式(9.5-11)我们看到互谱密度函数可以为复数，由于是将 $S_{XX}(\omega)$ 与一个任意的、一般的复数 H 相乘，因此互谱密度不再具有 $S_{XX}(\omega)$ 的非负的、偶对称的性质。而从式(9.5-14)看到，输出功率谱仍然具有与输入功率谱密度一样的实的、非负的偶函数的性质，这是由于与 $|H|^2$ 相乘不会改变这些性质。表9.5-2将以上提到的这些关系进行了归纳以备参考。

表9.5-2　线性系统在广义平稳随机过程输入下的输入/输出关系

WSS随机过程：	输出均值：
$Y(t)=h(t)*X(t)$	$\mu_Y=H(0)\mu_X$
互相关：	互功率谱密度：
$R_{XY}(\tau)=R_{XX}(\tau)*h^*(-\tau)$	$S_{XY}(\omega)=S_{XX}(\omega)H^*(\omega)$
$R_{YX}(\tau)=h(\tau)*R_{XX}(\tau)$	$S_{YX}(\omega)=H(\omega)S_{XX}(\omega)$
$R_{YY}(\tau)=R_{YX}(\tau)*h^*(-\tau)$	$S_{YY}(\omega)=S_{YX}(\omega)H^*(\omega)$
自相关：	功率谱密度：
$R_{YY}(\tau)=h(\tau)*R_{XX}(\tau)*h^*(-\tau)$	$S_{YY}(\omega)=\lvert H(\omega)\rvert^2S_{XX}(\omega)$
$=g(\tau)*R_{XX}(\tau)$	$=G(\omega)S_{XX}(\omega)$
输出功率及方差：	
$E\{\lvert Y[t]\rvert^2\}=R_{YY}(0)=\dfrac{1}{2\pi}\int_{-\infty}^{+\infty}\lvert H(\omega)\rvert^2S_{XX}(\omega)\mathrm{d}\omega$	
$\sigma_Y^2=R_{YY}[0]-\lvert\mu_Y\rvert^2$	

现在就可以功率谱密度清楚地解释为平均功率的密度与频率的关系。我们将会证明对所有 ω 都有 $S(\omega)\geqslant0$，在频带 (ω_1,ω_2) 内的平均功率可以通过对 $S(\omega)$ 在这个频带内积分得到。

定理9.5-1　设 $X(t)$ 是一个平稳的、二阶随机过程，其相关函数为 $R_{XX}(\tau)$，功率谱密度为 $S_{XX}(\omega)$。则 $S_{XX}(\omega)\geqslant0$ 并且对所有 $\omega_2\geqslant\omega_1$

$$\frac{1}{2\pi}\int_{\omega_1}^{\omega_2}S_{XX}(\omega)\mathrm{d}\omega$$

为频带 (ω_1,ω_2) 内的平均功率。

证明 设 $\omega_2 > \omega_1$，且都为实数。定义一个滤波器的传递函数为

$$H(\omega) \triangleq \begin{cases} 1, & \omega \in (\omega_1, \omega_2) \\ 0, & \text{其他} \end{cases}$$

注意到它只允许在 (ω_1, ω_2) 带内的信号通过。如果 $X(t)$ 输入到这个滤波器，则输出 $Y(t)$ 的功率谱密度为[由式(9.5-14)]

$$S_{YY}(\omega) = \begin{cases} S_{XX}(\omega), & \omega \in (\omega_1, \omega_2) \\ 0, & \text{其他} \end{cases}$$

现在输出 $Y(t)$ 的平均功率 $E[|Y(t)|^2] = R_{YY}(0)$

$$R_{YY}(0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{YY}(\omega)\mathrm{d}\omega = \frac{1}{2\pi}\int_{\omega_1}^{\omega_2} S_{XX}(\omega)\mathrm{d}\omega \geqslant 0$$

对所有 $\omega_2 > \omega_1$，上式都成立。则选择 $\omega_2 \approx \omega_1$，我们可以得到 ω，对所有 $S_{XX}(\omega) \geqslant 0$ 成立，函数 S_{XX} 被理解为功率谱密度是因为将其对一个频带积分，可以得到其在这个频带内的平均功率。

之前看到一个函数是一个有效的相关函数或协方差函数的条件是非常强的。事实上，我们看到这个函数必须是半正定的，尽管实际上并没有证明这个条件是充分的。我们看到在频域中进行分析的优势在于容易判断一个给定的频率函数是否为功率谱密度。这个函数必须是实的、非负的，也就是 $S(\omega) \geqslant 0$。对于一个给定的函数 $F(\omega) \geqslant 0$，假设功率谱密度为 $S_{WW} = 1$ 的白噪声通过一个传递函数为 $H(\omega) = \sqrt{F(\omega)}$ 的滤波器，则根据式(9.5-14)，输出的功率谱密度为 $S_{XX}(\omega) = F(\omega)$，这样就证明了 F 是一个有效的功率谱密度。如果随机过程为实值，绝大部分随机过程都满足，则还需要 $F(\omega)$ 为偶函数以满足定义 9.5-1 后列举的功率谱密度的性质(2)。所有这些可以总结如下：

定理 9.5-2 设 $F(\omega)$ 为一个实的非负的可积函数，即对所有 ω 有 $F(\omega) \geqslant 0$。那么，存在一个功率谱密度为 $S(\omega) = F(\omega)$ 的平稳随机过程。如果随机过程为实过程，则 $F(\omega)$ 一定是 ω 的偶函数。

我们看到要检验一个功率谱密度是否有效比证明相关函数的半正定性要容易些。事实上，很容易证明一个函数半正定性等价于它的傅里叶变换的非负性，因此半正定性是证明一个函数为有效的相关函数或协方差函数的充分条件。首先，由定理 9.5-2 知道，$S(\omega)$ 的非负性就包含了半正定的条件。为证明等效性，我们只剩下要证明函数 $f(\tau)$ 的半正定条件包含了其傅里叶变换 $F(\omega)$ 的非负性。继续证明如下：由于 $f(\tau)$ 是半正定的，则

$$\sum_{n=1}^{N}\sum_{m=1}^{N} a_n a_m^* f(\tau_n - \tau_m) \geqslant 0$$

而且由于

$$f(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(\omega)\mathrm{e}^{+\mathrm{j}\omega\tau}\mathrm{d}\omega$$

我们有

$$\frac{1}{2\pi}\sum_{n}\sum_{m}\left(a_n a_m^* \int_{-\infty}^{+\infty} F(\omega)\mathrm{e}^{+\mathrm{j}\omega(\tau_n-\tau_m)}\right)\mathrm{d}\omega \geqslant 0$$

这个式子可以写为

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}F(\omega)\left[\sum_n\sum_m a_n a_m^* \mathrm{e}^{+\mathrm{j}\omega(\tau_n-\tau_m)}\right]\mathrm{d}\omega=\frac{1}{2\pi}\int_{-\infty}^{+\infty}F(\omega)\left|\sum_{n=1}^{N}a_n\mathrm{e}^{+\mathrm{j}\omega\tau_n}\right|^2\mathrm{d}\omega\geqslant 0$$

我们看到在上式幅度平方项中的式子被称为横向或节拍延迟滤波器。这样只要选择的 N 足够大，且 τ_n 等间隔，就可以选择 a_n 的值来近似具有任意传递函数 $H(\omega)$ 的理想滤波器。通过将 H 设为以 ω 为中心的非常窄带的滤波器，就可以得到对所有 $\omega\,(-\infty<\omega<+\infty)$ 都有 $F(\omega)\geqslant 0$，于是得到了下一定理：

定理 9.5-3　$f(\tau)$ 是相关函数的充分必要条件是它必须为半正定的。

这里有一个与概率密度函数类似的情况，概率密度可以被看成其特征函数的傅里叶变换。我们已经知道，一个函数是有效的概率密度函数的充分条件是它必须为非负的（假定已被归一化为积分等于 1）；因此概率密度与功率谱密度类似；实际上，可以定义一个类似于累积分布函数的谱分布函数[9-7]。这样特征函数和相关函数也是类似的，所以它们为使其各自有效都必须为半正定的。此外概率密度的积分为 1 使得特征函数满足 $\Phi(0)=1$，这个条件很容易通过缩放一个不等于 0 的任意半正定函数来满足。

平稳过程和微分方程

我们现在来看统计微分方程，输入为一个平稳的或至少是广义平稳的随机过程，并且是一个任意时刻都有效的线性常系数微分方程（LCCDE）。假设方程在有界输入、有界输出条件下是稳定的，所以输出随机过程也是平稳的（或者广义平稳，如果输入也是广义平稳的）。

考虑以下 LCCDE：

$$\begin{aligned}&a_N Y^{(N)}(t)+a_{N-1}Y^{(N-1)}(t)+\cdots+a_0Y(t)\\&\quad=b_M X^{(M)}(t)+b_{M-1}X^{(M-1)}(t)+\cdots+b_0X(t),\qquad -\infty<t<+\infty\end{aligned}$$

这表明输出 $Y(t)$ 是输入 $X(t)$ 经过一个线性系统产生的，线性系统的频率响应为

$$H(\omega)=B(\omega)/A(\omega),\qquad \text{且}\, a_0\neq 0$$

其中

$$B(\omega)\triangleq\sum_{m=0}^{M}b_m(\mathrm{j}\omega)^m$$

和

$$A(\omega)\triangleq\sum_{n=0}^{N}a_n(\mathrm{j}\omega)^n$$

它是一个有理函数，分子 $B(\omega)$ 和分母 $A(\omega)$ 都为多项式。因为系统是稳定的，我们可以运用上一节的结论得到

$$\mu_Y=\mu_X H(0)$$

$$S_{YX}(\omega)=H(\omega)S_{XX}(\omega)$$

以及

$$S_{YY}(\omega)=|H(\omega)|^2S_{XX}(\omega)$$

其中

$$H(0)=b_0/a_0 \quad \text{和} \quad |H(\omega)|^2=|B(\omega)|^2/|A(\omega)|^2$$

因此

$$\mu_Y=(b_0/a_0)\,\mu_X \quad \text{和} \quad S_{YY}(\omega)=\left(|B(\omega)|^2/|A(\omega)|^2\right)S_{XX}(\omega)$$

这种频域分析方法一般会优于时域分析方法，但要求输入输出都至少是平稳随机过程。我们得到谱密度后，可以用 IFT 来得到相关函数和协方差函数。如果把 IFT 的计算过程看成双边的拉普拉斯逆变换，计算常常更容易。式(9.5-3)的拉普拉斯变换为

$$\boldsymbol{S}_{YX}(s)=\boldsymbol{H}(s)\boldsymbol{S}_{XX}(s) \tag{9.5-15}$$

而式(9.5-13)的拉普拉斯变换为

$$\boldsymbol{S}_{YY}(s)=\boldsymbol{H}(s)\boldsymbol{H}(-s)\boldsymbol{S}_{XX}(s) \tag{9.5-16}$$

因为 $\boldsymbol{h}^*(-\tau)\leftrightarrow\boldsymbol{H}(-s)$。回忆一下双边拉普拉斯变换的定义[9-3]，对任意 $f(\tau)$

$$\boldsymbol{F}(s)\triangleq\int_{-\infty}^{+\infty}f(\tau)\,\mathrm{e}^{-s\tau}\mathrm{d}\,\tau$$

注意到这样一个具有复变量 s 的函数可由具有实变量 ω 的傅里叶变换函数 $F(\omega)$ 得到，只需两个步骤，首先设

$$\boldsymbol{F}(s)_{s=\mathrm{j}\omega}\triangleq F(\omega)$$

然后用 s 取代 $\mathrm{j}\omega$。在第 8 章对于离散的情况我们用了类似的扩展方法，将傅里叶变换通过 Z 变换扩展到整个复平面。

例 9.5-6 （输出相关——一阶系统）考虑一阶微分方程

$$Y'(t)+\alpha Y(t)=X(t), \qquad \alpha>0$$

输入 $X(t)$ 平稳且均值 $\mu_X=0$，协方差函数 $K_{XX}(\tau)=\delta(\tau)$，则很容易的得到系统函数为

$$H(\omega)=\frac{1}{\alpha+\mathrm{j}\omega}$$

输入过程的功率谱密度为

$$S_{XX}(\omega)=1$$

于是可得到下列互功率谱密度和输出功率谱密度

$$S_{YX}(\omega)=H(\omega)S_{XX}(\omega)=\frac{1}{\alpha+\mathrm{j}\omega}$$

$$S_{YY}(\omega)=|H(\omega)|^2S_{XX}(\omega)=\frac{1}{|\alpha+\mathrm{j}\omega|^2}=\frac{1}{\alpha^2+\omega^2}$$

现在将其变为拉普拉斯变换，设 $s=\mathrm{j}\omega$

$$\mathsf{S}_{YY}(\mathrm{j}\omega)=\frac{1}{\left(\alpha^2-(\mathrm{j}\omega)^2\right)}$$

$$=\frac{1}{(\alpha+\mathrm{j}\omega)(\alpha-\mathrm{j}\omega)}$$

则

$$\boldsymbol{S}_{YY}(s)=\frac{1}{(s+\alpha)(-s+\alpha)}$$

用留数法(参见附录A)或部分分数展开法，可由拉普拉斯逆变换直接得到输出的自相关函数为

$$R_{YY}(\tau)=\frac{1}{2\alpha}\exp(-\alpha|\tau|),\qquad -\infty<\tau<+\infty$$

由于 $\mu_Y=0$，所以它也是输出的协方差函数。通过上面 $S_{XY}(\omega)$ 方程，还可以得到互相关函数为 $R_{YX}(\tau)=\exp(-\alpha\tau)u(\tau)$。

在例9.5-6中很有趣的一点是对于 $\tau<0$ 有 $R_{YX}(\tau)=0$。这意味着在输入 X 为白噪声的情况下输出 Y 与输入 X 的所有未来值正交。之所以会这样有两个原因：系统是因果系统和输入为白噪声。系统的因果性要求输出不直接依赖于未来的输入值，而是只依赖于现在和过去的输入值。输入的白噪声特性确保了过去和现在的输入值与未来的输入值。将这两个条件联系起来，我们看到当前的输出与未来的输入之间没有互相关性。如果进一步假设输入是高斯的，则输入过程是一个独立过程，输出与所有未来的输入独立。则我们可以说系统的因果性阻止了当前时刻的输出当未来输入值的直接依赖性，独立的输入过程阻止了任何直接的依赖性。这个概念对用于估计理论的马尔可夫过程(参见第11章)的理论是非常重要的。

例9.5-7 (输出相关函数——二阶系统)考虑以下二阶LCCDE系统

$$\frac{\mathrm{d}^2Y(t)}{\mathrm{d}t^2}+3\frac{\mathrm{d}Y(t)}{\mathrm{d}t}+2Y(t)=5X(t)$$

与上一个例子类似，仍然是白噪声输入。这里系统函数是

$$H(\omega)=\frac{5}{(\mathrm{j}\omega)^2+3\mathrm{j}\omega+2}=\frac{5}{(2-\omega^2)+\mathrm{j}3\omega}$$

则与例9.5-6类似，输出功率谱密度为

$$S_{YY}(\omega)=\frac{25}{(2-\omega^2)^2+(3\omega)^2}=\frac{25}{\omega^4+5\omega^2+4}$$

运用留数法计算IFT，我们定义一个复变量的函数 $\boldsymbol{S}_{YY}(s)|_{s=\mathrm{j}\omega}\triangleq S_{YY}(\omega)$，用复变量 $\mathrm{j}\omega$ 重写右边可得到

$$\mathsf{S}_{YY}(\mathrm{j}\omega)=\frac{25}{(\mathrm{j}\omega)^4-5(\mathrm{j}\omega)^2+4}$$

代入 $s=\mathrm{j}\omega$，可得到

$$\boldsymbol{S}_{YY}(s)=\frac{25}{s^4-5s^2+4}$$

做因式分解

$$\frac{5}{(s+2)(s+1)}\cdot\frac{5}{(-s+2)(-s+1)}=\boldsymbol{H}(s)\boldsymbol{H}(-s)$$

其中 $\boldsymbol{H}(s)$ 是拉普拉斯变换系统函数。然后由拉普拉斯逆变换得到输出的相关函数

$$R_{YY}(\tau)=25\left[\frac{1}{6}\exp(-|\tau|)-\frac{1}{12}\exp(-2|\tau|)\right],\qquad -\infty<\tau<+\infty$$

我们将详细的计算过程留给感兴趣的读者。

9.6 周期和循环平稳过程

除了平稳和广义平稳随机过程，还有两类其他的随机过程经常会遇到。它们就是周期和循环平稳随机过程，这里给出它们的定义：

定义 9.6-1　一个随机过程 $X(t)$ 是广义周期的，如果存在一个 $T>0$，对所有的 t 有

$$\mu_X(t)=\mu_X(t+T)$$

并且对所有的 t_1，t_2 有

$$K_{XX}(t_1,t_2)=K_{XX}(t_1+T,t_2)=K_{XX}(t_1,t_2+T)$$

最小的 T 称为周期，注意到 $K_{XX}(t_1,t_2)$ 是以 T 为周期的周期函数。

一个广义周期随机过程的例子是例 9.4-1 中提到的随机复指数过程。实际上，过程的随机傅里叶级数表示为

$$X(t)=\sum_{k=1}^{\infty}A_k\exp\left(\mathrm{j}\frac{2\pi kt}{T}\right) \tag{9.6-1}$$

其中的随机变量系数 A_k 也是广义周期的。广义周期过程也可能是 WSS，这种情况我们称其为广义周期平稳。这类过程将在网上的第 10 章进一步讨论，我们称为均方周期。一个广义周期过程的协方差函数在图 9.6-1 中勾画出来，可看到 $K_{XX}(t_1,t_2)$ 是二维周期函数，其周期为 (T,T)。在网上的第 10 章会看到，除了某些概率为零的事件集，广义周期随机过程的样本函数依概率 1 为周期函数，即

$$X(t)=X(t+T)$$

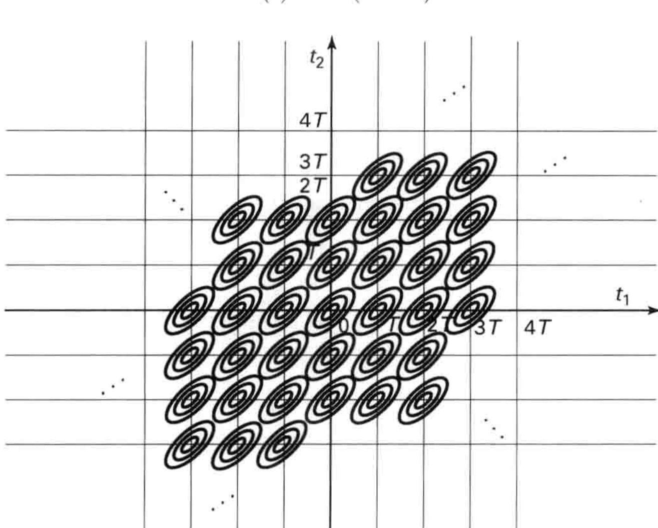


图 9.6-1　广义(WS)周期随机过程的协方差函数可能的轮廓图

另一个重要的分类是循环平稳。它只是部分与周期性相关的，但非常容易混淆。读者要仔细区分以下定义的不同。大体上说，循环平稳过程具有周期的统计特性，但周期过程的样本函数是周期的。

定义9.6-2 一个随机过程$X(t)$是广义循环平稳的，如果存在一个正的T满足对所有t有

$$\mu_X(t)=\mu_X(t+T)$$

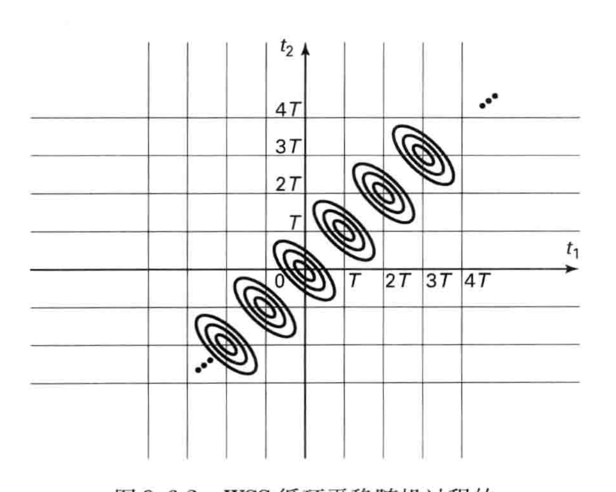
并且对所有t_1，t_2有

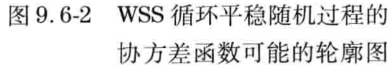
$$K_{XX}(t_1,t_2)=K_{XX}(t_1+T,t_2+T)$$

一个循环平稳的例子是式(9.2-11)中的随机PSK过程。它的均值函数为0，因此是一般周期的。它的协方差函数[参见式(9.2-13)]对两个变量的平移是不变的。注意式(9.2-13)不是双周期的[因为$R_{XX}(0,\ T)\ =0\neq R_{XX}(0,\ 0)$]。还需要注意$X(t)$的样本函数在任何意义下都不是周期的。

一个典型的循环平稳随机过程的协方差函数的常数值轮廓在图9.6-2中画出。注意这张图与图9.6-1中显示的周期随机过程的区别。实际上，循环平稳意味着其统计特性是具有周期性的，但随机过程本身不是周期的。

沿着45°线(即$t_1=t_2$)求平均，可以得到这两类过程的广义平稳随机过程。图9.6-3中显示了这个周期过程恒定密度的轮廓变为了WSS周期过程的直线形式。这个循环平稳过程的WSS过程刚好变成了一般的WSS过程，因为在135°线($t_1=-t_2$)上缺少任何的周期结构。




图9.6-2 WSS循环平稳随机过程的协方差函数可能的轮廓图

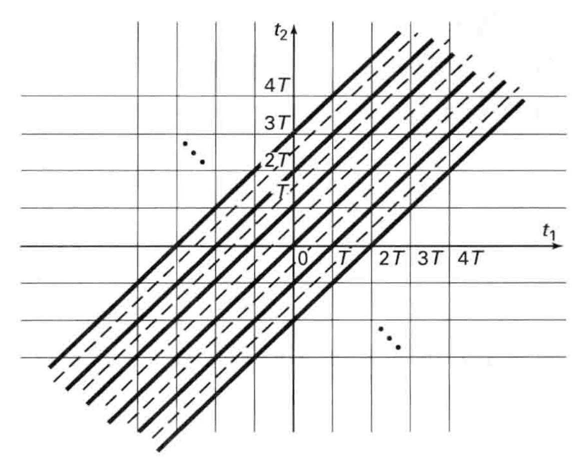


图9.6-3 WSS周期随机过程的协方差函数可能的轮廓图(实线是最大值，虚线是最小值)

除了调制器，扫描传感器也会产生循环平稳过程。例如，在电视机中的逐行扫描器就将随机图像场变成了一个可用循环平稳随机过程建模的一维随机过程。在通信系统中，循环平稳经常会因为波形以波特率或符号率重复而出现。

在信号处理中，当一个平稳随机序列被一个滤波器组和二次采样分析时可能出现循环平稳。接下来的滤波器组综合包括上采样和重构滤波器。如果二次采样的周期是N，则产生的综合随机序列会是周期为N的循环平稳过程。当用完美的重构滤波器时，综合的输出会达到真实的平稳。

当循环平稳过程不是平稳或非WSS时，有时将一个循环平稳过程转化为一个平稳过程是合适的，其过程如下所示。

例9.6-1 (WSS PSK)在9.2节中，我们已经看到PSK过程是循环平稳的。因此不是WSS，这由式(9.2-13)很容易看出。循环平稳的出现是基于这样的事实：模拟角过程$\Theta_a(t)$

是分段常数的，而且只在 $t=nT$(n 为整数)时变化。在许多实际情况下，调制过程开始于任意时刻 t，从系统设计者的角度考虑，这实际上可以用一个随机量来建模。因此在这一实际情况中，通过增加一个随机变量 T_0，调制信号过程[参见方程(9.2-11)]转换为

$$\widetilde{X}(t)=\cos(2\pi f_c t+\Theta_a(t)+2\pi f_c T_0) \tag{9.6-2}$$

T_0 在[0, T]之间均匀分布，并且与角过程 $\Theta_a(t)$ 独立。则我们很容易看出，均值和协方差函数只需要对 T_0 求集平均来修正即可，而由于 T_0 是均匀分布的，所以只需在[0, T]进行积分即可。因此可得到

$$\begin{aligned} R_{\tilde{X}\tilde{X}}(t_1+\tau,t_1) &= \frac{1}{T}\int_0^T R_{XX}(t_1+\tau+t,t_1+t)\mathrm{d}t \\ &= \frac{1}{T}\int_{-\infty}^{+\infty} s_Q(t_1+t+\tau)s_Q(t_1+t)\mathrm{d}t \\ &= \frac{1}{T}s_Q(\tau)*s_Q(-\tau) \end{aligned} \tag{9.6-3}$$

刚好是位移 τ 的函数，则 $\tilde{X}(t)$ 是一个 WSS 随机过程。

例 9.6-2 (PSK 的功率谱密度)在例 9.6-1 中通过一个平均过程，我们定义一个随机 PSK 信号的 WSS 形式，其中平均是对消息时间或波特间隔 T 内进行的。得到的 WSS 随机过程 $\tilde{X}(t)$ 的相关函数[参见方程(9.6-3)]是

$$R_{\tilde{X}\tilde{X}}(\tau)=\frac{1}{T}s_Q(\tau)*s_Q(-\tau)$$

其中 $s_Q(\tau)$ 为

$$s_Q(\tau)=\begin{cases}\sin(2\pi f_c\tau), & 0\leqslant\tau\leqslant T\\ 0, & \text{其他}\end{cases}$$

则这个 PSK 信号的 WSS 形式的功率谱密度为

$$\begin{aligned} S_{\tilde{X}\tilde{X}}(\omega) &= FT\{R_{\tilde{X}\tilde{X}}(\tau)\} \\ &= \frac{1}{T}|FT\{s_Q(\tau)\}|^2 \\ &\approx (T/4)\left\{\left(\frac{\sin(\omega-2\pi f_c)\frac{T}{2}}{(\omega-2\pi f_c)\frac{T}{2}}\right)^2+\left(\frac{\sin(\omega+2\pi f_c)\frac{T}{2}}{(\omega+2\pi f_c)\frac{T}{2}}\right)^2\right\} \\ & f_cT>>1 \end{aligned} \tag{9.6-4}$$

其图形可用 MATLAB 绘制①。文件 psd_ PSK.m 在本书的网站上可查阅。

一些图是用 psd_ PSK.m 画的，针对两组不同的 f_c 和 T 的值。首先来看图 9.6-4 中的 psd 图，当 $f_c=2.5$，$T=0.5$ 时，$S_{\tilde{X}\tilde{X}}(\omega)$ 的正频率和负频率部分有相当大的重叠。在载频 f_c 处没有集中的功率是不足为奇的，因为在波特间隔 T 内只包括了 $s_Q(t)$ 的一个周期很少的部分。接下来的一对图则显示了完全不同的情况，功率高度集中在 ω_c 处。这张图是在 $f_c=3.0$，$T=5.0$ 条件下计算得到的，因此在波特间隔 T 内有正弦波形的 15 个周期。图 9.6-5 是一个线性图，而图 9.6-6 是 $S_{\tilde{X}\tilde{X}}(\omega)$ 对数图。

---

① 近似等号的原因是，我们忽略了式(9.6-4)中两个 sinc 信号在 $\pm f_c$ 处的交叉项，因为 $f_cT>>1$。

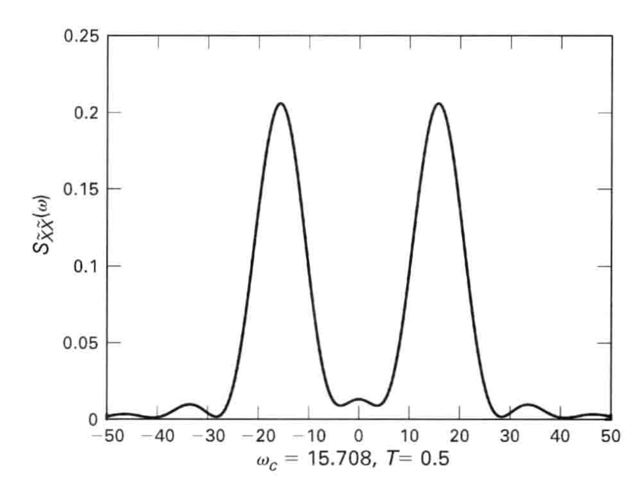


图9.6-4 PSK的功率谱密度，其中f_c = 2.5，T = 0.5

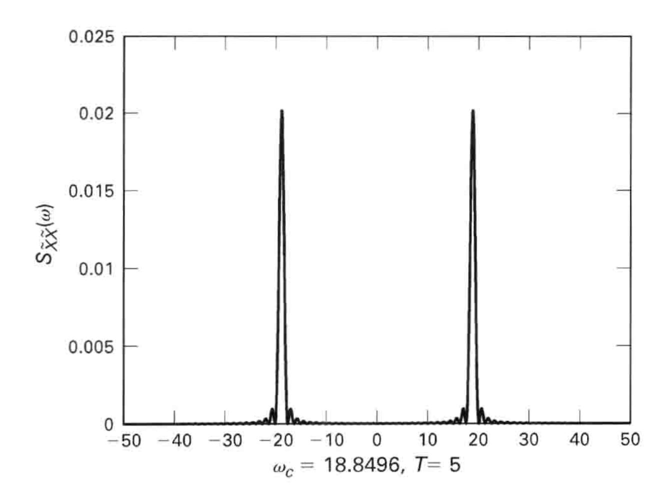


图9.6-5 PSK的功率谱密度，其中f_c = 3，T = 5

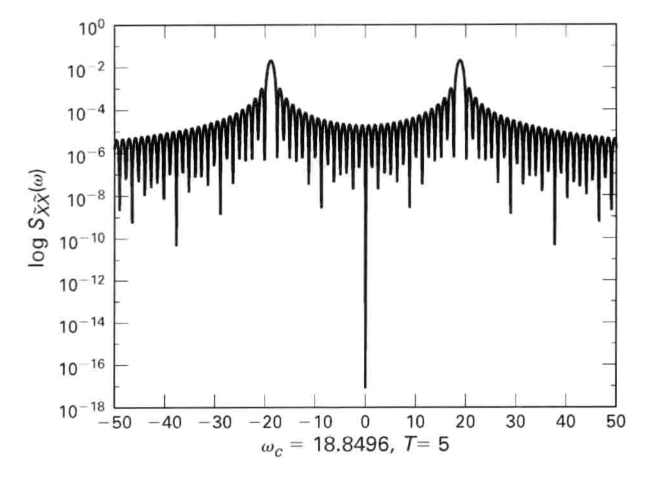


图9.6-6 PSK的功率谱密度的对数图，其中f_c = 3，T = 5

9.7 向量过程和状态方程

在这一节中我们会将9.5节中的一些结论推广到重要的向量随机过程类中。因此会带来一个关于状态方程和向量马尔可夫过程的简短讨论。向量随机过程会在通信系统带通信号的同相和正交分量的双通道系统中出现。向量过程也广泛地应用于控制系统用来对具有多输入和输出的工业过程建模。同时，向量模型也会从高阶标量模型里人为地创造出来，为了运用在估计和控制理论中非常有用的“状态”概念。

设 $X_1(t)$ 和 $X_2(t)$ 是两个联合平稳随机过程，分别输入到系统 H_1 和 H_2 中，输出为 Y_1 和 Y_2，如图9.7-1所示。

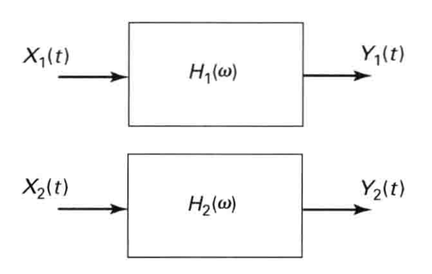


图9.7-1 一个通用的(无关的)两通道LSI系统

由之前的讨论知道如何计算 $R_{X_1Y_1}$，$R_{X_2Y_2}$，$R_{Y_1Y_1}$，$R_{Y_2Y_2}$。我们现在来看看怎么计算系统之间的相关函数，也就是 $R_{X_1Y_2}$，$R_{X_2Y_1}$ 和 $R_{Y_1Y_2}$。给定 $R_{X_1X_2}$，首先计算

$$\begin{aligned}R_{X_1Y_2}(\tau) &= E[X_1(t+\tau)Y_2^*(t)] \\ &= \int_{-\infty}^{+\infty} E[X_1(t+\tau)X_2^*(t-\beta)]h_2^*(\beta)\mathrm{d}\beta \\ &= \int_{-\infty}^{+\infty} R_{X_1X_2}(\tau+\beta)h_2^*(\beta)\mathrm{d}\beta \\ &= \int_{-\infty}^{+\infty} R_{X_1X_2}(\tau-\beta')h_2^*(-\beta')\mathrm{d}\beta', \qquad (\beta'=-\beta)\end{aligned}$$

于是

$$R_{X_1Y_2}(\tau) = R_{X_1X_2}(\tau) * h_2^*(-\tau)$$

通过对称性得到

$$R_{X_2Y_1}(\tau) = R_{X_2X_1}(\tau) * h_1^*(-\tau)$$

输出的互相关为

$$R_{Y_1Y_2}(\tau) = h_1(\tau) * R_{X_1X_2}(\tau) * h_2^*(-\tau)$$

结果用功率谱表示为

$$S_{X_1Y_2}(\omega) = S_{X_1X_2}(\omega)H_2^*(\omega)$$

以及

$$S_{Y_1Y_2}(\omega) = H_1(\omega)H_2^*(\omega)S_{X_1X_2}(\omega)$$

我们注意以下重要的事实：如果两个系统函数 H_1 和 H_2 的支集①不重叠，则输出 Y_1 和 Y_2 是正交随机过程并且与输入过程中任何相关函数无关。可以用图9.7-2所示的带内部耦合的双通道系统来表示。这里有两个附加的系统函数来表示输入和输出的交叉耦合。它们用 H_{12} 和 H_{21} 来表示。

这种情况最好用向量法处理；因此我们定义

① 回忆一下函数 g 的支集定义为

$$\mathrm{supp}(g) \triangleq \{x|g(x)\neq 0\}$$

$$\boldsymbol{X}(t) \triangleq [X_1(t), X_2(t)]^{\mathrm{T}}, \qquad \boldsymbol{Y}(t) \triangleq [Y_1(t), Y_2(t)]^{\mathrm{T}}$$

以及

$$\boldsymbol{h}(t) \triangleq \begin{bmatrix} h_{11}(t) & h_{12}(t) \\ h_{21}(t) & h_{22}(t) \end{bmatrix}$$

其中 $h_{ij}(t)$ 是具有频率响应 $H_{ij}(\omega)$ 的子系统的冲激响应。于是得到

$$\boldsymbol{Y}(t) = \boldsymbol{h}(t) * \boldsymbol{X}(t) \tag{9.7-1}$$

其中向量卷积定义为

$$(\boldsymbol{h}(t) * \boldsymbol{X}(t))_i \triangleq \sum_{j=1}^{N} h_{ij}(t) * X_j(t)$$

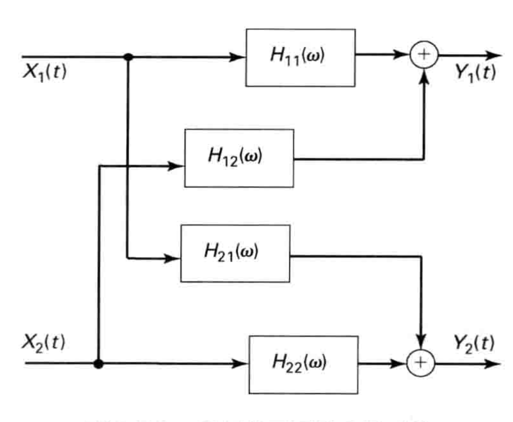


图 9.7-2　通用的两通道 LSI 系统

如果定义以下的输入输出的相关矩阵

$$\boldsymbol{R}_{\boldsymbol{XX}}(\tau) \triangleq \begin{bmatrix} R_{X_1X_1}(\tau) & R_{X_1X_2}(\tau) \\ R_{X_2X_1}(\tau) & R_{X_2X_2}(\tau) \end{bmatrix} \tag{9.7-2}$$

$$\boldsymbol{R}_{\boldsymbol{YY}}(\tau) \triangleq \begin{bmatrix} R_{Y_1Y_1}(\tau) & R_{Y_1Y_2}(\tau) \\ R_{Y_2Y_1}(\tau) & R_{Y_2Y_2}(\tau) \end{bmatrix} \tag{9.7-3}$$

则可以证明(参见习题 9.44)

$$\boldsymbol{R}_{\boldsymbol{YY}}(\tau) = \boldsymbol{h}(\tau) * \boldsymbol{R}_{\boldsymbol{XX}}(\tau) * \boldsymbol{h}^{\dagger}(-\tau) \tag{9.7-4}$$

其中的 † 代表埃尔米特(或共轭)转置。

进行矩阵的傅里叶变换，可得到

$$\boldsymbol{S}_{\boldsymbol{YY}}(\omega) = \boldsymbol{H}(\omega)\boldsymbol{S}_{\boldsymbol{XX}}(\omega)\boldsymbol{H}^{\dagger}(\omega) \tag{9.7-5}$$

其中

$$\boldsymbol{H}(\omega) = FT\{\boldsymbol{h}(t)\}$$

$$\boldsymbol{S}(\omega) = FT\{\boldsymbol{R}(\tau)\}$$

在这个符号中意味着隐含了按矩阵分量依次求傅里叶变换。这种多通道的情况显然是通过扩大相关矩阵的维数来使其应用到 M 个输入和 N 输出的情况中。

状态方程

正如习题 9.43 所证明的那样，可以将 N 阶 LCCDE 重写为一阶向量微分方程的形式，其中输出向量的维数等于 N

$$\dot{\boldsymbol{Y}}(t) = \boldsymbol{A}\boldsymbol{Y}(t) + \boldsymbol{B}\boldsymbol{X}(t), \qquad -\infty < t < +\infty \tag{9.7-6}$$

这刚好是式(9.7-1)的多通道系统，也可以被解释为一组 N 个耦合的一阶 LCCDE。运用向量傅里叶变换，计算得到系统函数为

$$\boldsymbol{H}(\omega) = (\mathrm{j}\omega\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B} \tag{9.7-7}$$

其中 $\boldsymbol{I}$ 是单位阵。此外，也可用卷积形式表示这一运算为

$$\boldsymbol{Y}(t) = \boldsymbol{h}(t) * \boldsymbol{X}(t)$$

在这里假设多通道系统是稳定的；也就是，所有的冲激响应 h_{ij} 是 BIBO 稳定的。接下来的求解过程就与一阶方程的标量求解过程大致相同；实际上，可以证明

$$\boldsymbol{h}(t)=\exp(\boldsymbol{A}t)\boldsymbol{B}u(t) \tag{9.7-8}$$

在本章前面部分对一个连续时间马尔可夫链的概率向量的求解过程中曾遇到过矩阵指数函数 $\exp(\boldsymbol{A}t)$。这个函数被广泛应用于线性系统理论，在参考文献[9-3]中已详细研究过其性质。

如果计算 WSS 情况下的互相关矩阵，可得到

$$\boldsymbol{R}_{\boldsymbol{YX}}(\tau)=\exp(\boldsymbol{A}\tau)\boldsymbol{B}u(\tau)*\boldsymbol{R}_{\boldsymbol{XX}}(\tau)$$

和

$$\boldsymbol{R}_{\boldsymbol{XY}}(\tau)=\boldsymbol{R}_{\boldsymbol{XX}}(\tau)*\boldsymbol{B}^{\dagger}\exp(-\boldsymbol{A}^{\dagger}\tau)u(-\tau)$$

输出相关矩阵为

$$\boldsymbol{R}_{\boldsymbol{YY}}(\tau)=\boldsymbol{h}(\tau)*\boldsymbol{R}_{\boldsymbol{XX}}(\tau)*\boldsymbol{h}^{\dagger}(-\tau) \tag{9.7-9}$$

做向量傅里叶变换，则有

$$\boldsymbol{S}_{\boldsymbol{YY}}(\omega)=(\mathrm{j}\omega\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B}\boldsymbol{S}_{\boldsymbol{XX}}(\omega)\boldsymbol{B}^{\dagger}(-\mathrm{j}\omega\boldsymbol{I}-\boldsymbol{A}^{\dagger})^{-1} \tag{9.7-10}$$

如果 $\boldsymbol{R}_{\boldsymbol{XX}}(\tau)=\boldsymbol{Q}\delta(\tau)$，则由于系统 $\boldsymbol{H}$ 假定是因果的，也就是，当 $t<0$ 时，$\boldsymbol{h}(t)=0$，我们得到当 $\tau<0$ 时互相关矩阵 $\boldsymbol{R}_{\boldsymbol{YX}}(\tau)=\boldsymbol{0}$；也就是当 $\tau<0$ 时 $E[\boldsymbol{Y}(t+\tau)\boldsymbol{X}^{*}(t)]=\boldsymbol{0}$。因此 $\boldsymbol{Y}(t)$ 的过去值与 $\boldsymbol{X}(t)$ 的当前和未来值都是正交的。如果进一步假设输入过程 $\boldsymbol{X}(t)$ 是一个高斯过程，那么不相关变为独立条件。则在高斯假设下，输出 $\boldsymbol{Y}(t)$ 与 $\boldsymbol{X}(t)$ 现在和未来独立。类似的结果在对标量情况进行分析时已经得到。可以利用这个结果证明，一阶向量 LCCDE 的类是一个具有如下定义的向量马尔可夫随机过程。

定义 9.7-1 （向量马尔可夫）随机过程 $\boldsymbol{Y}(t)$ 是一个马尔可夫过程，只要满足对所有 $n>0$ 和 $t_n>t_{n-1}>\cdots>t_1$ 以及所有 $\boldsymbol{y}(t_{n-1}),\cdots,\boldsymbol{y}(t_1)$ 的值，有

$$P[\boldsymbol{Y}(t_n)\leqslant\boldsymbol{y}_n|\boldsymbol{y}(t_{n-1}),\cdots,\boldsymbol{y}(t_1)]=P[\boldsymbol{Y}(t_n)\leqslant\boldsymbol{y}_n|\boldsymbol{y}(t_{n-1})]$$

对所有实数向量 $\boldsymbol{y}_n$ 成立。这里 $\boldsymbol{A}\leqslant\boldsymbol{a}$ 的含义是

$$(A_n\leqslant a_n,A_{n-1}\leqslant a_{n-1},\cdots,A_1\leqslant a_1)$$

在讨论向量微分方程之前，我们先简要回顾一下确定性向量 LCCDE 系统的一个结论。一阶向量方程为

$$\dot{\boldsymbol{y}}(t)=\boldsymbol{A}\boldsymbol{y}(t)+\boldsymbol{B}\boldsymbol{x}(t),\qquad t\geqslant t_0$$

给定初始条件 $\boldsymbol{y}(t_0)$，利用矩阵指数可以证明

$$\boldsymbol{y}(t)=\exp[\boldsymbol{A}(t-t_0)]\boldsymbol{y}(t_0)+\int_{t_0}^{t}\boldsymbol{h}(t-v)\boldsymbol{x}(v)\mathrm{d}v,\qquad t\geqslant t_0$$

这样推广了标量情况。在线性系统理论中可以找到这种确定性的解，例如参考文献[9-3]。第一项称为零输入解，第二项称为零状态(或驱动)解，解类似于标量 LCCDE。

我们可以将这个理论扩展到随机情况，只需要在半无限域 $t_0\leqslant t<\infty$ 考虑微分方程(9.7-6)，并用下面随机解替换确定性解即可

$$\boldsymbol{Y}(t)=\exp[\boldsymbol{A}(t-t_0)]\boldsymbol{Y}(t_0)+\int_{t_0}^{t}\boldsymbol{h}(t-v)\boldsymbol{X}(v)\mathrm{d}v \tag{9.7-11}$$

如果 LCCDE 是 BIBO 平稳的，也就是，$\boldsymbol{A}$ 的特征值的实部都为负数，则当 $t_0\to-\infty$ 时，我们得到了对于所有时刻的解为

$$\boldsymbol{Y}(t)=\int_{-\infty}^{t}\boldsymbol{h}(t-v)\boldsymbol{X}(v)\mathrm{d}v=\boldsymbol{h}(t)*\boldsymbol{X}(t) \tag{9.7-12}$$

这与我们已在平稳的无限时间域内推导出的结论相同。实际上，利用系统的稳定性就可以得出在任何有限的时间内这个解的零输入部分一定是零。

下面的定理将证明用上述方法生成一个向量高斯-马尔可夫随机过程的方法。现在输入是白高斯向量过程 $\boldsymbol{W}(t)$，输出向量是马尔可夫过程，用 $\boldsymbol{X}(t)$ 表示。

定理 9.7-1 设状态方程为

$$\dot{\boldsymbol{X}}(t)=\boldsymbol{A}\boldsymbol{X}(t)+\boldsymbol{B}\boldsymbol{W}(t)$$

输入为白高斯过程 $\boldsymbol{W}(t)$，则输出 $\boldsymbol{X}(t)$ 为一个向量高斯-马尔可夫随机过程。

证明 用 t_{n-1} 时刻的解表示 t_n 时刻的解，即

$$\boldsymbol{X}(t_n)=\exp[\boldsymbol{A}(t_n-t_{n-1})]\boldsymbol{X}(t_{n-1})+\int_{t_{n-1}}^{t_n}\boldsymbol{h}(t_n-v)\boldsymbol{W}(v)\mathrm{d}v$$

然后将积分表示为 $\boldsymbol{I}(t_n)$，注意它与 $\boldsymbol{X}(t_{n-1})$ 是独立的。于是可以推出

$$\begin{aligned}&P[\boldsymbol{X}(t_n)\leqslant\boldsymbol{x}_n|\boldsymbol{x}(t_{n-1}),\cdots,\boldsymbol{x}(t_1)]\\&=P[\boldsymbol{I}(t_n)\leqslant\boldsymbol{x}_n-\mathrm{e}^{\boldsymbol{A}(t_n-t_{n-1})}\boldsymbol{x}(t_{n-1})|\boldsymbol{x}(t_{n-1}),\cdots,\boldsymbol{x}(t_1)]\\&=P[\boldsymbol{I}(t_n)\leqslant\boldsymbol{x}_n-\mathrm{e}^{\boldsymbol{A}(t_n-t_{n-1})}\boldsymbol{x}(t_{n-1})|\boldsymbol{x}(t_{n-1})]\end{aligned}$$

因此 $\boldsymbol{X}(t)$ 是一个向量马尔可夫过程。

如果在定理 9.7-1 没有对输入 $\boldsymbol{W}(t)$ 做高斯假设，而只是白噪声，则得不到输出是马尔可夫的结论。因为没有在证明中用到独立性条件，而只有较弱的不相关条件。另一方面，如果放松高斯条件的约束而要求输入 $\boldsymbol{W}(t)$ 是一个独立的随机过程，那么过程 $\boldsymbol{X}(t)$ 仍然是马尔可夫的，但不是高斯-马尔可夫的。我们在这个定理中用 $\boldsymbol{X}$ 而不是 $\boldsymbol{Y}$ 是强调 LCCDE 通常是用来对输入过程建模的。

小结

在这一章中介绍了随机过程的概念，它是一个含连续参数的函数集合。参数一般是时间，但也可以是位置或另一个连续的变量。本章的大部分内容讨论的是扩展到二维和三维参数。事实上，许多现代的应用需要一个二维参数，例如，图像的灰度函数 $i(t_1, t_2)$。这些随机函数被称为随机场，可以利用本章方法的扩展来进行分析。随机场在参考文献[9-5]和参考文献[9-8]的第 7 章和许多其他地方都有讨论。

我们介绍了一些重要的过程：异步二进制信号，泊松计数过程，随机电报信号和相移键控，它们是数字通信的基础；维纳过程，第一个高斯随机过程的例子，是非线性滤波理论的基本构建模块；马尔可夫过程，由于它的有效性和易于处理使其得到了广泛的应用，是网上第 11 章中卡尔曼-布西滤波器中采用的信号模型。

考虑了线性系统对随机过程的二阶特性的影响。我们将结果运用于平稳和 WSS 过程，介绍了功率谱密度和 LSI 系统的对应分析。我们还简要地考虑了广义周期和循环平稳过程，引入了随机向量过程和系统并将马尔可夫模型扩展应用到了其中。

习题

（带 * 号的习题有一定的难度，需要花费更多精力对教材进一步的学习。）

9.1 设 $X[n]$ 是实平稳随机序列，均值 $E\{X[n]\}=\mu_X$，自相关函数 $E\{X[n+m]X[n]\}=R_{XX}[m]$。如果

$X[n]$是一个 D/A 转换器的输入，则其连续时间输出信号可被理想化为模拟随机过程 $X_a(t)$

$$X_a(t) \triangleq X[n], \qquad n \leqslant t < n+1 \text{ ，对所有 } n$$

如图 P9.1 所示。

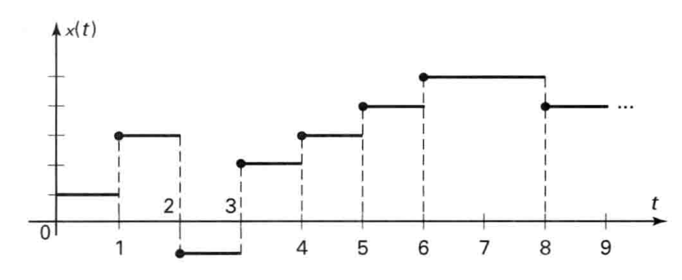


图 P9.1 采样-保持 D/A 转换器的典型输出

(a)求均值 $E[X_a(t)] = \mu_a(t)$，表示为 μ_X 的函数。

(b)求相关函数 $E[X_a(t_1)X_a(t_2)] = R_{X_aX_a}(t_1, t_2)$，用 $R_{XX}[m]$表示。

9.2 考虑一个 WSS 随机序列 $X[n]$，均值 μ_X 为常数，自相关函数为 $R_{XX}[m]$，构造一个随机过程

$$X(t) \triangleq \sum_{n=-\infty}^{+\infty} X[n]\frac{\sin \pi(t-nT)/T}{\pi(t-nT)/T}, \quad -\infty < t < +\infty$$

接下来，假设无限和是收敛的，无须担心统计收敛问题。

(a)求均值 $\mu_X(t)$，用 μ_X 表示，尽可能地简化结论。

(b)求相关函数 $R_{XX}(t_1, t_2)$，用 $R_{XX}[m]$表示。$X(t)$是否广义平稳？

提示：线性系统的采样原理表明任意带宽有限的确定函数 $g(t)$可由其均匀的空间样本完全恢复出来，即

$$g(t) = \sum_{n=-\infty}^{+\infty} g(nT)\frac{\sin \pi(t-nT)/T}{\pi(t-nT)/T}$$

只要函数 $g(t)$的带宽小于或等于 π/T。

9.3 令 $B[n]$是一个伯努利随机序列，等可能取 ±1。然后定义一个随机过程 $X(t) \triangleq \sqrt{p}\sin(2\pi f_o t + B[n]\frac{\pi}{2})$

$nT \leqslant t < (n+1)T$，对所有 n，其中$\sqrt{p}$和f_0 是实数。

(a)求均值函数。

(b)求协方差函数。

9.4 $Y(t)$是如图 P9.4 所示的节拍延迟线滤波器的输出信号，输入为 $X(t)$，N 为节拍数

$$Y(t) = \sum_{n=0}^{N-1} A_n X(t-nT)$$

输入 $X(t)$是一个平稳高斯随机过程，均值为 0，自相关函数为 $R_{XX}(\tau)$，且具有性质 $R_{XX}(nT) = 0$ 对所有 $n \neq 0$ 的整数都成立。节拍增益 $A_n(n = 0.1, \cdots, N-1)$也是零均值、不相关的高斯随机变量，方差为 σ_A^2。每个节拍增益都与输入过程 $X(t)$独立。

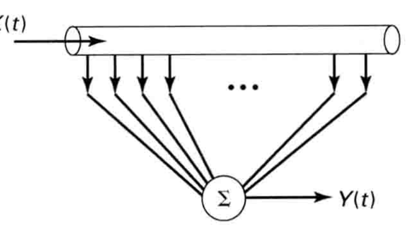


图 P9.4 节拍延迟线滤波器

(a)求 $Y(t)$的自相关函数。

(b)对一个给定值 t，求 $Y(t)$的特征函数。解释具体的过程。

(c)对固定时刻 t，求当 $N \to \infty$ 时，$\frac{1}{\sqrt{N}}Y(t)$的渐进 pdf，请解释。

(d)假设节拍数 N 是一个泊松随机变量，均值为 $\lambda(>0)$，再求(a)和(b)。

(注意：可能需要公式 $e^{-x}\approx\frac{1}{1+x}$，对所有 $|x|\ll 1$ 和 $e^x=\sum_{n=0}^{\infty}\frac{x^n}{n!}$。)

9.5 设 $N(t)$ 是一个定义在 $0\leqslant t<\infty$ 上的泊松随机过程，$N(0)=0$，平均到达率 $\lambda>0$。

(a)求联合概率 $P[N(t_1)=n_1, N(t_2)=n_2]$，其中 $t_2>t_1$。

(b)求 K 阶联合 PMF 的表达式

$$P_N(n_1,\cdots,n_K;t_1,\cdots,t_K)$$

其中 $0\leqslant t_1<t_2<\cdots<t_K<\infty$，注意考虑 n_1，…，n_K 的有关值。

*9.6 一个在 $t\geqslant 0$ 上的非均匀泊松计数过程 $N(t)$ 定义如下：

(a) $N(0)=0$。

(b) $N(t)$ 具有独立增量。

(c)对所有 $t_2\geqslant t_1$

$$P[N(t_2)-N(t_1)=n]=\frac{\left[\int_{t_1}^{t_2}\lambda(v)\mathrm{d}v\right]^n}{n!}\exp\left(-\int_{t_1}^{t_2}\lambda(v)\mathrm{d}v\right),\qquad n\geqslant 0$$

函数 $\lambda(t)$ 被称为强度函数，并且对所有 t 有 $\lambda(t)\geqslant 0$，即始终非负。

(a)求这个非均匀分布的泊松过程的均值 $\mu_N(t)$。

(b)求 $N(t)$ 的相关函数 $R_{NN}(t_1, t_2)$。定义时间轴的一个变换如下：

$$\tau(t)\triangleq\int_0^t\lambda(v)\mathrm{d}v$$

对所有 v，如果 $\lambda(v)>0$，则 $\tau(t)$ 是单调递增的，因此我们可以定义逆映射 $t(\tau)$，如图 P9.6 所示。

(c)对所有 t 假设 $\lambda(t)>0$，定义计数过程

$$N_u(\tau)\triangleq N(t(\tau))$$

证明 $N_u(\tau)$ 是一个均匀泊松计数过程，其率参数 $\lambda=1$；即证明对 $\tau\geqslant 0$：

(1) $N_u(0)=0$。

(2) $N_u(\tau)$ 有独立增量。

(3)对所有 $\tau_2\geqslant\tau_1$

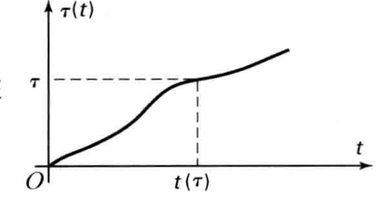


图 P9.6 τ 随 t 的变化图

$$P[N_u(\tau_2)-N_u(\tau_1)=n]=\frac{(\tau_2-\tau_1)^n}{n!}e^{-(\tau_2-\tau_1)},\qquad n\geqslant 0$$

9.7 一个非均匀泊松过程 $N(t)$ 具有强度函数(平均到达率)

$$\lambda(t)=1+2t$$

对所有 $t\geqslant 0$，初始值 $N(0)=0$。

(a)求均值函数 $\mu_N(t)$。

(b)求相关函数 $R_{NN}(t_1, t_2)$。

(c)求 $N(t)\geqslant t$ 的概率表达式，即对任意 $t>0$，求 $P[N(t)\geqslant t]$。

(d)求(c)用错误函数 $\mathrm{erf}(x)$ 表示的近似解。

*9.8 这个问题关注的是 9.2 节中给出的有关泊松计数过程的构建。

(a)证明第 n 个到达时间 $T[n]$ 的密度为

$$f_T(t;n)=\frac{\lambda^n t^{n-1}}{(n-1)!}e^{-\lambda t}u(t),\qquad n>0$$

在泊松过程的增量是泊松分布性质的推导过程中，即

$$P[X(t_a)-X(t_b)=n]=\frac{[\lambda(t_a-t_b)]^n}{n!}e^{-\lambda(t_a-t_b)}u[n],\qquad t_a>t_b$$

我们隐含地利用了这样一个事实，在$(t_b, t_a]$上第一个到达时距是服从指数分布的。实际上，这一事实并不清晰，因为问题中提到的到达时距只是部分在区间$(t_b, t_a]$，图 P9.8 显示了一个示意图。定义$\tau'[i] \triangleq T[i]-t_b$为部分到达时距，我们注意到$\tau'[i]=\tau[i]-T$，其中随机变量$T \triangleq t_b - T[i-1]$，$\tau[i]$表示了到达时距。

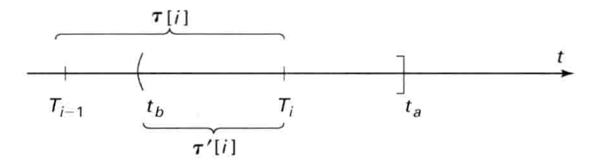


图 P9.8 在任意观测间隔内到达时间的关系图解

(b)固定随机变量$T=t$，求 CDF：

$$F_{\tau'[i]}(\tau'|T=t) = P\{\tau[i] \leqslant \tau' + t|\tau[i] \geqslant t\}$$

(c)将(b)中条件修改成设T为随机变量，求τ'的无条件 CDF(提示：这部分的内容不需要很多的计算)。

由于前面提到的性质，指数分布被称为无记忆的。它是唯一具有这个性质的连续分布。

9.9 设$N(t)$为$[0, \infty)$上一个计数过程，其平均率参数$\lambda(t)$依赖于另一个正随机过程$S(t)$，特殊情况为$\lambda(t)=S(t)$。假设给定$\{S(t)$在$[0, \infty)\}$上的$N(t)$为非均匀泊松过程。已知$\mu_S(t)=\mu_0>0$和$K_{SS}(t_1, t_2)$。

(a)求$t \geqslant 0$的$\mu_N(t)$，用μ_0表示。

(b)求$t \geqslant 0$的$\sigma_N^2(t)$，用$K_{SS}(t_1, t_2)$表示。

9.10 设随机过程$K(t)$依赖于一个平均到达率为$\lambda>0$的均匀泊松过程$N(t)$，可描述为：在$t=0$时刻起始，$N(t)=0$且$K(t)=0$。当一个到达在$N(t)$发生，一个独立的伯努利试验以成功概率p发生，其中$0<p<1$。如果成功，$K(t)$加1，否则$K(t)$不变化。具体的过程参见图 P9.11。求离散随机过程$K(t)$在时刻t的一阶 PMF，即$t \geqslant 0$的$P_K(k; t)$。

9.11 设图像的扫描线可由空间随机过程$S(x)$描述，用$S(x)$对点x处的理想灰度级建模。通过调整光源的强度，用光通道独立地发射每个点

$$\lambda(t,x) = S(x) + \lambda_0, \qquad 0 \leqslant t \leqslant T$$

用这种方法，我们得到了一簇随机过程，由连续参数x索引

$$\{N(t,x)\}$$

对每个x，$N(t, x)$在给定$S(x)$时是均匀泊松过程(参见图 P9.11)。在观察结束时，存储$N(x) \triangleq N(T, x)$，求这个空间过程的统计特性。

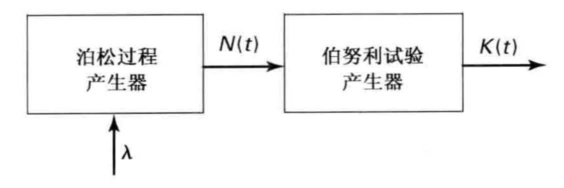


图 P9.11 泊松调制的伯努利试验过程

总结如下：$N(x)$是一个整数值的空间随机过程，它依赖于另一个被称为信号过程的随机过程$S(x)$的值。这个空间随机过程$S(x)$是零均值的平稳过程，协方差函数为

$$K_{SS}(x) = \sigma_S^2 \exp(-\alpha|x|)$$

其中$\alpha>0$。$N(x)$在给定$S(x)=s(x)$的条件分布为泊松分布，均值为$\lambda(x)=(s(x)+\lambda_0)T$，其中$\lambda_0$为正常数，即

$$P[N(x)=n|S(x)=s(x)]=\frac{\lambda^n(x)}{n!}\mathrm{e}^{-\lambda(x)}u[n]$$

随机变量$N(x)$从点到点是条件独立的。

(a)求(无条件的)均值和方差

$$\mu_N(x)=E[N(x)] \quad 和 \quad E\left[(N(x)-\mu_N(x))^2\right]$$

(提示：首先求出条件均值和条件均方值。)

(b)求$R_{NN}(x_1, x_2)\triangleq E[N(x_1)N(x_2)]$。

9.12 设$X(t)$为定义在$t\geqslant 0$上的随机电报信号(RTS)。固定$X(0)=+1$。RTS用一个泊松随机到达时间序列$T[n]$在± 1间切换$X(t)$的值。设平均到达率为λ(大于0)，则有

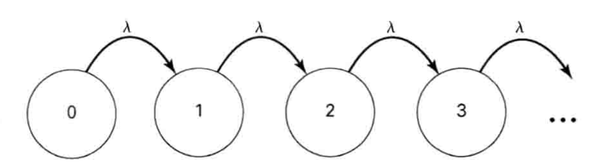
$$X(t)\triangleq\begin{cases}1, & 0\leqslant t<T[1]\\ -1, & T[1]\leqslant t<T[2]\\ +1, & T[2]\leqslant t<T[3]\\ \cdots, & \cdots\end{cases}$$

(a)证明$X(t)$为一个马尔可夫过程，画出其状态转移图。

(b)求$X(t)=+1$的稳态概率，即$P_X(1;\infty)$，用率参数λ来表示。

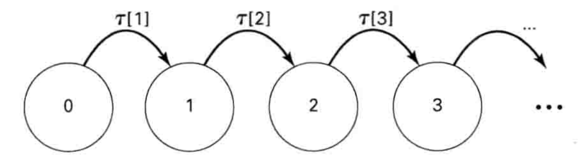
(c)求状态概率$P_X(1;t)$和$P_X(-1;t)$的微分方程。

*9.13 一个率参数为λ(大于0)的均匀泊松过程$N(t)$为一个无限状态的马尔可夫链，其状态转移图如图P9.13(a)所示。其中状态标签就是过程$N(t)$的值。而且独立的到达时间$\tau[n]$是服从参数为λ的指数分布。



图P9.13(a) 用马尔可夫链表示的泊松过程

我们对上面的情景做以下改动。用任意非负的、平稳的、独立的随机序列取代独立的到达时距$\tau[n]$，仍用$\tau[n]$表示，生成了一个通常称为更新过程的随机过程，如图P9.13(b)所示。



图P9.13(b) 更一般的(更新)过程链

(a)证明更新过程的PMF $P_N(n;t)=P[N(t)=n]$可用到达时间的CDF $F_T(t;n)$表示为

$$P_N(n;t)=F_T(t;n)-F_T(t;n+1), \qquad n\geqslant 1$$

其中到达时间$T[n]=\sum_{k=1}^{n}\tau[k]$，$F_T(t;n)$为到达时间$T[n]$的CDF。

(b)设$\tau[n]$服从分布$U[0,1]$，即在$[0,1]$上服从均匀分布，求这个特殊的更新过程在$n=0$，1和2时刻的$P_N(n;t)$。

(c)求到达时间序列$T[n]$的特征函数。

(d)求(b)中更新过程的CDF $F_T(t;n)$近似表达式，n越大，表达式越精确，且与$T[n]$均值差距越小(提示：当x较小时，有$\sin x\approx x-x^3/3!$成立)。

9.14 令$W(t)$为定义在$[0,\infty)$上的维纳过程[即t时刻的分布为$N(0,t)$]。求联合密度$f_W(a_1,a_2;t_1,t_2)$，$0<t_1<t_2$。

9.15 设 $W_1(t)$ 和 $W_2(t)$ 是两个维纳过程，相互独立，都定义在 $t \geqslant 0$ 上，方差参数分别为 α_1 和 α_2。设过程 $X(t)$ 定义为其代数差函数，即 $X(t) \triangleq W_1(t) - W_2(t)$。

(a)求当 $t_1, t_2 \geqslant 0$ 时 $R_{XX}(t_1, t_2)$。

(b)求当 $t \geqslant 0$ 时 $X(t)$ 的 pdf $f_X(x; t)$。

*9.16 设随机过程 $X(t)$ 具有均值函数 $\mu_X(t) = 4$，协方差函数 $K_{XX}(t_1,t_2) = 5[\min(t_1,t_2)]^2$。设导数过程记为 $Y(t) = X'(t)$，$t \geqslant 0$。

(a)求 $Y(t)$ 的均值函数。

(b)求相关函数 $R_{YY}(t_1,t_2)$。

(c)导数过程 $Y(t)$ 是否为广义平稳的？

(d)证明过程 $X(t)$ 可对标准维纳过程构建。

*9.17 设 $W(t)$ 是一个标准维纳过程，即 $\alpha = 1$，定义

$$X(t) \triangleq W^2(t), \qquad t \geqslant 0$$

(a)求概率密度 $f_X(x; t)$。

(b)求条件概率密度 $f_X(x_2 \mid x_1; t_2, t_1)$，$t_2 > t_1$。

(c) $X(t)$ 是马尔可夫过程吗？为什么？

(d) $X(t)$ 是否具有独立增量？请证明。

9.18 设 $X(t)$ 是定义在 $[0, \infty)$ 上的马尔可夫随机过程，初始概率密度 $f_X(x; 0) = \delta(x-1)$，条件 pdf 为

$$f_X(x_2|x_1;t_2,t_1) = \frac{1}{\sqrt{2\pi(t_2-t_1)}}\exp\left(-\frac{1}{2}\frac{(x_2-x_1)^2}{t_2-t_1}\right), \quad \text{对于所有} t_2 > t_1$$

(a)求 $f_X(x; t)$，t 为任意值。

(b)设 $f_X(x; 0) \sim N(0, 1)$，求 $f_X(x, t)$。

9.19 考虑三状态马尔可夫链 $N(t)$，其状态转移图如图 P9.19 所示。其中状态标签都是实际的输出，例如当链处在状态 3 时，即 $N(t) = 3$。状态之间的转移都是联合独立且服从指数分布的到达时距所控制，其平均速率在图中分枝上显示。

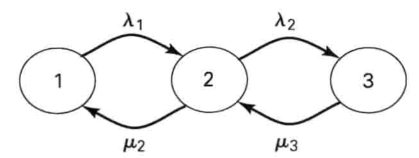


图 P9.19 一个三状态连续时间马尔可夫链

(a)假设 $t = 0$ 时刻从状态 2 起始，求对任意 $t > 0$ 满足直到 t 时刻仍处在这个状态的概率为多少(提示：有两种方法可离开状态 2。因此你必须有比两个独立的指数随机变量更少的时间离开状态 2，这两个指数随机变量的率参数分别为 μ_2 和 λ_2)。

(b)写出求 $t \geqslant 0$ 时刻处在状态 i 的概率的微分方程，表示为 $p_i(t)$，$i = 1, 2, 3$[提示：首先写出用 $p_i(t)$ 表示的 $p_i(t + \delta t)$，$i = 1, 2, 3$，仅保留 $O(\delta t)$ 项]。

(c)求 $p_i(t)$ 当 $i = 1, 2, 3$ 时的稳态解，即 $p_i(\infty)$。

9.20 设某无线通信二进制信道处于好的或坏的状态可用图 P9.20 中描述的具有某转移率的连续时间马尔可夫链来描述。假定呈指数分布状态转移率的率参数为 $\lambda_1 = 1$，$\lambda_2 = 9$。每一种状态 ε 的值在(b)中给定。

(a)求信道处于好的状态的稳态概率。设 $P\{X(t) = \text{good}\} = p_G$，$P\{X(t) = \text{bad}\} = p_B$(提示：假设稳态存在，然后根据 $t - \delta$ 时刻的两种可能性写出 t 时刻的 p_G，只保留一阶 δ，设 δ 很小)。

(b)假设处在好的状态时，二进制信道中无错误，但处在坏的状态时错误概率 $\varepsilon = 0.01$，求信道的平均错误概率(假设信道在每个单字节传输过程中不会发生状态改变)。

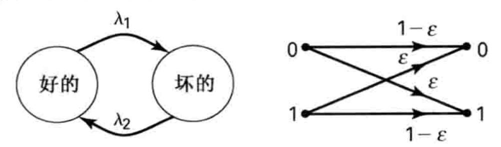


图 P9.20 两状态无线通信信道的模型

9.21　这个问题关注对具有连续振幅的马尔可夫过程 $X(t)$ 的查普曼-科尔莫戈罗夫方程[参见式(9.2-22)]对于三个增加的观测时刻 $t_3 > t_2 > t_1 > 0$ 的条件 pdf 为

$$f_X(x(t_3)|x(t_1)) = \int_{-\infty}^{+\infty} f_X(x(t_3)|x(t_2))f_X(x(t_2)|x(t_1))\,\mathrm{d}x(t_2)$$

你将证明协方差函数为 $K_{XX}(t,s) = \alpha\min(t,s)\,(\alpha > 0)$ 的维纳过程，解以上方程。

(a)求 $t>0$ 时这个维纳过程的一阶 pdf $f_X(x(t))$。

(b)求 $t>s>0$ 时一阶条件 pdf $f_X(x(t) \mid x(s))$。

(c)参考查普曼-科尔莫戈罗夫方程，设 $t_3 - t_2 = t_2 - t_1 = \delta$，用 x_3，x_2，x_1 表示对应时刻的值。然后证明(b)中的条件 pdf 满足以下方程：

$$f_X(x_3|x_1) = \int_{-\infty}^{+\infty} f_X(x_3|x_2)f_X(x_2|x_1)\,\mathrm{d}x_2$$

9.22　例 9.3-2 中的随机过程 $X'(t)$ 是否平稳？为什么？

9.23　设 A 和 B 为独立同分布的随机变量，均值为 0，方差为 σ^2，三阶矩 $m_3 \triangleq E[A^3] = E[B^3] \neq 0$，考虑以下随机过程

$$X(t) = A\cos(2\pi ft) + B\sin(2\pi ft), \qquad -\infty < t < +\infty$$

其中 f 为一给定频率。

(a)证明随机过程 $X(t)$ 是 WSS。

(b)证明 $X(t)$ 不是严平稳。

9.24　前面我们证明定理 9.3-2，推导了式(9.3-5)。描述和证明协方差函数的对应定理。并不假定 $\mu_X(t) = 0$。

9.25　假定 $x(t)$ 为平稳随机过程，均值为 u_X，协方差为 $k_{XX}(\tau) = \delta(\tau)$。假定 $x(t)$ 的样本函数驱动以下微分方程：

$$\dot{Y}(t) + aY(t) = X(t), \qquad a>0,\ -\infty < t < +\infty$$

(a) 求 $\mu_Y(t) = E[Y(t)]$。

(b) 求 $R_{YY}(\tau)$。

(c) 求 $\sigma_Y^2(t)$。

9.26　在习题 9.1 中由 D/A 转换器产生的随机过程 $X(t)$ 是 WSS 吗？

9.27　随机过程的功率谱为 $S_{XX}(\omega) = 1/(\omega^2+9)$，$-\infty < \omega < +\infty$，求它的自相关函数 $R_{XX}(\tau)$。

9.28　考虑如图 P9.28 所示一个 LSI 系统，其输入为零均值随机过程 $W(t)$，输出为随机过程 $X(t)$。系统的频率响应为 $H(\omega)$。给定 $K_{WW}(\tau) = \delta(\tau)$，求用互协方差函数 $K_{XW}(\tau)$ 或其傅里叶变换表示 $H(\omega)$。

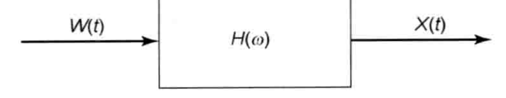


图 P9.28　白噪声输入下的 LSI 系统

9.29　假定随机过程 $Y(t)$ 由下式给出：

$$Y(t) = X(t) + 0.3\frac{\mathrm{d}X(t)}{\mathrm{d}t}, \qquad -\infty < t < +\infty$$

其中 $X(t)$ 是随机过程，其均质函数为 $u_X(t) = 5t$ 和协方差函数为

$$K_{XX}(t_1, t_2) = \frac{\sigma^2}{1+\alpha(t_1-t_2)^2}, \qquad \alpha > 0$$

(a) 求均值函数 $\mu_Y(t)$。

(b) 求协方差函数 $K_{YY}(t_1, t_2)$。

(a) $Y(t)$ 是 WSS 吗？为什么？

9.30　考虑一阶随机微分方程

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} + X(t) = W(t)$$

由零均值、自相关函数为 $R_{WW}(t,s) = \delta(t-s)$ 的白噪声 $W(t)$ 驱动。

(a)如果微分方程对所有时间 $-\infty < t < \infty$ 都有效，求输出的广义平稳随机过程 $X(t)$ 的 psd。

(b)用留数定理(或其他方法)求 $S_{XX}(\omega)$ 的傅里叶逆变换，即自相关函数 $R_{XX}(\tau)$，$-\infty < \tau < \infty$。

(c)如果以上微分方程只对 $t>0$ 成立，是否可以找到一个初始条件随机变量 $X(0)$ 使得 $X(t)$ 对所有 $t>0$都为广义平稳的？如果这样一个随机变量存在，求其均值和方差。你可以假设这个随机变量 $X(0)$ 与 $W(t)$ 在 $t \geqslant 0$ 时正交，即 $X(0) \perp W(t)$ [提示：$t>0$ 时的 $X(t)$ 可以根据初始条件和含有 $W(t)$ 的随机积分来表示]。

9.31 证明 $h(\tau) * h^*(-\tau)$ 是半正交函数。假定函数 $h(t)$ 是平方可积函数，即 $\int_{-\infty}^{+\infty} |h()t|^2 \mathrm{d}t < \infty$。

9.32 设随机过程 $X(t)$，均值为 128，协方差函数

$$K_{XX}(\tau) = 1000\exp(-10|\tau|)$$

用一个低通滤波器滤波

$$H(\omega) = \frac{1}{1 + \mathrm{j}\omega}$$

产生输出过程当 $Y(t)$。

(a)求均值函数 $\mu_Y(t)$。

(b)求协方差函数 $K_{YY}(\tau)$。

9.33 考虑连续时间系统，输入随机过程为 $X(t)$，输出过程为 $Y(t)$

$$Y(t) = \frac{1}{4}\int_{-2}^{+2} X(t-s)\mathrm{d}s$$

假设输入 $X(t)$ 是 WSS，功率谱密度 $S_{XX}(\omega) = 2$，$-\infty < \omega < +\infty$。

(a)求输出功率谱密度 $S_{YY}(\omega)$。

(b)求输出的相关函数 $R_{YY}(\tau)$。

9.34 一个零均值的 WSS 随机过程 $Y(t)$ 的样本函数如图 P9.34 所示，它是具有随机振幅和随机持续时长连续的矩形脉冲。

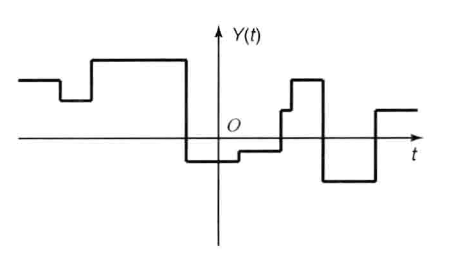


图 P9.34 随机振幅脉冲列

脉冲宽度的概率密度为

$$f_W(w) = \begin{cases} \lambda e^{-\lambda w}, & w \geqslant 0 \\ 0, & w < 0 \end{cases}$$

其中 $\lambda > 0$。每个脉冲的振幅是一个随机变量 X(与 W 独立)，其均值为 0，方差为 σ_X^2。各振幅和脉宽是独立的。

(a)求自相关函数 $R_{YY}(\tau) = E[Y(t+\tau)Y(t)]$。

(b)求对应的功率谱密度 $S_{YY}(\omega)$。

[提示：首先求出条件自相关函数 $E[Y(t+\tau)Y(t) | W = w]$，其中 t 假设为脉冲的起始(这样做不丢失 $Y(t)$ WSS 假设的一般性)]。

9.35 设 $X(t)$ 为 WSS 随机过程，均值为 μ_X，协方差函数为 $K_{XX}(\tau) = 5\cos(10\tau)e^{-2|\tau|}$。过程 $X(t)$ 输入到一线性系统，系统函数为

$$H(s) = \frac{s+4}{s^2 + 11s + 30}$$

产生输出过程 $Y(t)$。

(a)首先，求输入功率谱密度 $S_{XX}(\omega)$。画出草图。

(b)写出 $X(t)$ 在 $\omega_1 \leqslant |\omega| \leqslant \omega_2$ 频率范围内的平均功率的表达式。可以积分形式表示。

(c)求输出功率谱密度 $S_{YY}(\omega)$。画出草图。

*9.36 在本题中我们考虑用白噪声的近似(更多关于白噪声的问题参见 9.5 节)，加到低通滤波器，通过对滤波器的输出过程的分析来考察用白噪声近似带来的误差。设零均值平稳随机过程 $X(t)$ 的自协方差函数为

$$K_{XX}(\tau)=\frac{1}{2\tau_0}\exp(-|\tau|/\tau_0)$$

上式可写为 $h(\tau)*h(-\tau)$，其中 $h(\tau)=\frac{1}{\tau_0}\mathrm{e}^{-\tau/\tau_0}u(\tau)$。

(a) 设 $X(t)$ 为低通滤波器的输入，如图 P9.36(a) 所示，输出为 $Y(t)$。求当 $G(\omega)$ 为下式所示时，输出 $Y(t)$ 的功率谱 $S_Y(\omega)$

$$G(\omega)\triangleq\begin{cases}1, & |\omega|\leqslant\omega_0\\ 0, & \text{其他}\end{cases}$$

(b) 直接用零均值的标准白噪声 $W(t)$ 作为系统的激励，白噪声的协方差函数为 $K_{WW}(\tau)=\delta(\tau)$。如图 P9.36(b) 所示，输出称为 $V(t)$，求输出的功率谱密度 $S_{VV}(\omega)$。

(c) 证明对于 $|\omega_0\tau_0|\ll 1$，$S_{YY}\approx S_{VV}$，并求功率差值 $|R_{VV}(0)-R_{YY}(0)|$ 的上限。



图 P9.36 (a) 对白噪声的近似输入到滤波器　　图 P9.36 (b) 白噪声输入到滤波器

9.37 考虑如图 P9.37 所示的 LSI 系统，设 $X(t)$ 和 $N(t)$ 为 WSS，且不相关，功率谱密度分别为 $S_{XX}(\omega)$ 和 $S_{NN}(\omega)$，均值为 0。

(a) 求输出 $Y(t)$ 的功率谱密度。

(b) 求 X 和 Y 的互谱密度，即求 $S_{XY}(\omega)$ 和 $S_{YX}(\omega)$。

(c) 定义误差 $\xi(t)\triangleq Y(t)-X(t)$，计算 $\xi(t)$ 的功率谱密度。

(d) 设 $h(t)=a\delta(t)$，选择一个 a 值使 $E[\xi^2(t)]=R_{\xi\xi}(0)$ 最小。

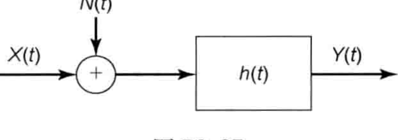


图 P9.37

9.38 设随机过程 $X(t)$ 定义为

$$X(t)\triangleq N\cos(2\pi f_0 t+\Theta)$$

其中 f_0 为一个已知频率，N 和 Θ 是独立的随机变量。N 的特征函数为

$$\Phi_N(\omega)=E[\mathrm{e}^{+\mathrm{j}\omega N}]=\exp\{\lambda[\mathrm{e}^{\mathrm{j}\omega}-1]\}$$

其中 λ 是一给定的正常数（即 N 是一个泊松随机变量）。随机变量 Θ 在 $[-\pi,+\pi]$ 间均匀分布。

(a) 确定均值函数 $\mu_X(t)$。

(b) 确定协方差函数 $K_{XX}(t,s)$。

(c) $X(t)$ 是否为 WSS？证明你的结论。

(d) $X(t)$ 是否平稳？证明你的结论。

9.39 设 $X(t)$ 为定义在 $t\geqslant 0$ 上的独立增量过程，初值为 $X(0)=X_0$ 是一个随机变量。假设下列特征函数存在：

$$E[\mathrm{e}^{\mathrm{j}\omega X_0}]\triangleq\Phi_{X_0}(\omega)\ \text{和}\ E[\mathrm{e}^{\mathrm{j}\omega(X(t)-X_0(s))}]\triangleq\Phi_{X(t)-X_0(s)}(\omega),\qquad t\geqslant s$$

(a) 根据定义 $E[\mathrm{e}^{\mathrm{j}\omega X(t)}]\triangleq\Phi_{X(t)}(\omega)$，证明

$$\Phi_{X(t)}(\omega)=\Phi_{X_0}(\omega)\Phi_{X(t)-X_0}(\omega)$$

(b) 证明对所有 $t_2\geqslant t_1$，$X(t_2)$ 和 $X(t_1)$ 的联合特征函数为

$$\Phi_{X(t_2),X(t_1)}(\omega_2,\omega_1)=\Phi_{X_0}(\omega_1+\omega_2)\Phi_{X(t_1)-X_0}(\omega_1+\omega_2)\Phi_{X(t_2)-X(t_1)}(\omega_2)$$

(c) 用高斯特征函数将 (a) 的结论运用到习题 9.18(b) 中。

9.40 用 pdf 的形式来表示以下问题的答案。

(a) 阐述一个独立增量随机过程的定义。

(b)阐述一个马尔可夫随机过程的定义。

(c)证明任意具有独立增量的随机过程也具有马尔可夫性。

9.41 设 $X(t)$ 为定义在 $t \geqslant 0$ 上的随机过程，具有独立增量，均值为 $\mu_X(t)=\mu_0$，协方差函数为

$$K_{XX}(t_1,t_2)=\sigma_X^2(\min(t_1,t_2))$$

其中 $\sigma_X^2(t)$ 为一个递增函数，即对所有 $t \geqslant 0$ 有 $\mathrm{d}\sigma_X^2(t)/\mathrm{d}t>0$，称为方差函数，注意 $\mathrm{Var}[X(t)]=\sigma_X^2(t)$。给定 $T>0$，求对所有 $t \geqslant T$，$Y(t) \triangleq X(t)-X(T)$ 的均值和协方差函数(注意：求协方差函数时取 t_1 和 t_2 均大于 T)。

*9.42 继续例 9.2-3，用 MATLAB 计算一个 $\alpha=2$ 和 $T=0.01$ 的维纳过程 $X(t)$ 的 1000 个元素的样本函数。

(a)用 MATLAB 程序 `hist.m` 来计算 $X(10)$ 的直方图并将它与理想的高斯 pdf 进行比较。

(b)用 `mean.m` 估计 $X(10)$ 的均值，`std.m` 估计它的标准差，并与理论值比较[提示：在一个 `for` 循环中用 `Wiener.m`① 来计算 `x(1000)` 的 100 个现实。然后用 `hist`。问题：为什么不能只用向量 `x` 的最后 100 个值来估计来近似得到要求的统计特性]。

9.43 设 WSS 随机过程 $X(t)$ 为以下三阶微分方程的输入：

$$\frac{\mathrm{d}^3Y}{\mathrm{d}t^3}+a_2\frac{\mathrm{d}^2Y}{\mathrm{d}t^2}+a_1\frac{\mathrm{d}Y}{\mathrm{d}t}+a_0Y(t)=X(t)$$

输出为 WSS 随机过程 $Y(t)$。

(a)将这个方程表示为以下的一阶向量微分方程形式：

$$\frac{\mathrm{d}\boldsymbol{Y}}{\mathrm{d}t}=\boldsymbol{AY}(t)+\boldsymbol{BX}(t)$$

定义 $\boldsymbol{Y}(t) \triangleq \begin{bmatrix} Y(t) \\ Y'(t) \\ Y''(t) \end{bmatrix}$，$\boldsymbol{X}(t) \triangleq [X(t)]$，求矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的值。

(b)求用输入 $\boldsymbol{R}_{\boldsymbol{XX}}(\tau)$ 表示 $\boldsymbol{R}_{\boldsymbol{XY}}(\tau)$ 的一阶矩阵微分方程。

(c)求用输入 $\boldsymbol{R}_{\boldsymbol{XY}}(\tau)$ 表示 $\boldsymbol{R}_{\boldsymbol{YY}}(\tau)$ 的一阶矩阵微分方程。

(d)用矩阵傅里叶变换证明输出的功率谱密度矩阵 $\boldsymbol{S}_{\boldsymbol{YY}}$ 为

$$\boldsymbol{S}_{\boldsymbol{YY}}(\omega)=(\mathrm{j}\omega\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{BS}_{\boldsymbol{XX}}(\omega)\boldsymbol{B}^{\dagger}(-\mathrm{j}\omega\boldsymbol{I}-\boldsymbol{A}^{\dagger})^{-1}$$

9.44 设 $\boldsymbol{X}(t)$ 为一个 WSS 向量随机过程，输入到 LSI 系统，系统的冲激响应矩阵为 $\boldsymbol{h}(t)$。

(a)证明输出 $\boldsymbol{Y}(t)$ 的相关矩阵为式(9.7-4)。

(b)推导矩阵协方差函数的对应方程。

9.45 在地球物理学信号处理中经常需要仿真一个多通道的随机过程。接下来的问题引出了一个关于这种向量随机过程的功率谱密度矩阵的重要约束。设 N 维向量随机过程 $\boldsymbol{X}(t)$ 为 WSS，相关矩阵为

$$\boldsymbol{R}_{\boldsymbol{XX}}(\tau) \triangleq E[\boldsymbol{X}(t+\tau)\boldsymbol{X}^{\dagger}(t)]$$

功率谱密度矩阵为

$$\boldsymbol{S}_{\boldsymbol{XX}}(\omega) \triangleq \mathrm{FT}\{\boldsymbol{R}_{\boldsymbol{XX}}(\tau)\}$$

FT{·}定义了矩阵傅里叶变换，即 $\boldsymbol{S}_{\boldsymbol{XX}}$ 的第(i,j)个元素是由 $\boldsymbol{R}_{\boldsymbol{XX}}$ 的第(i,j)个元素(即 $E[X_i(t+\tau)X_j^*(t)]$，其中 $X_i(t)$ 为 $X(t)$ 的第 i 个元素)傅里叶变换得到的。

(a)用常数 $a_1,\cdots,a_N$ 定义一个 WSS 标量过程

$$Y(t) \triangleq \sum_{i=1}^{N} a_i X_i(t)$$

求 $Y(t)$ 的功率谱密度，用矩阵 $\boldsymbol{S}_{\boldsymbol{XX}}(\omega)$ 的元素来表示。

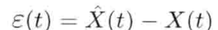
① `Wiener.m` 在本书的网站上提供。

(b)证明功率谱密度矩阵 $\boldsymbol{S}_{XX}(\omega)$ 对各确定的 ω 一定为一半正定矩阵，即对所有复列向量 $\boldsymbol{a}$ 一定有 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{S}_{XX}(\omega)\boldsymbol{a}^{*} \geqslant 0$。

9.46　考虑图 P9.46 所示线性系统，系统由两零均值、正交、联合 WSS 的"信号"$X(t)$和"噪声"$U(t)$激励，则系统 G 的输入为

$$Y(t) = h(t) * X(t) + U(t)$$

这可用来为在噪声中畸变的信号估计问题建模。如果将这个"接收的信号"$Y(t)$通过滤波器 G，可得到估计值 $\hat{X}(t)$，则 $\varepsilon(t)$ 可看成"估计误差"



$$\varepsilon(t) = \hat{X}(t) - X(t)$$

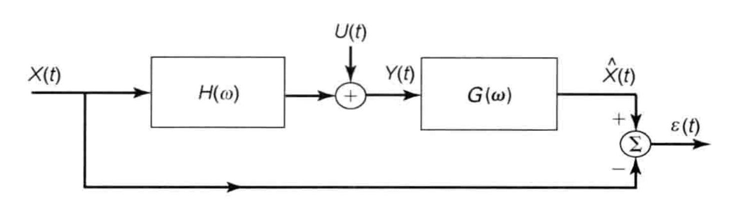


图 P9.46　用来评估估计误差的系统

在本题中我们将计算有关的功率谱密度以及互谱密度。

(a)求 $S_{YY}(\omega)$。

(b)求 $S_{X\hat{X}}(\omega) = S_{\hat{X}X}^{*}(\omega)$，用 H, G, S_{XX} 和 S_{UU} 表示。

(c)求 $S_{\varepsilon\varepsilon}(\omega)$。

(d)将你的答案用到(c)中证明。为了要使 $S_{\varepsilon\varepsilon}(\omega)$ 在

$$S_{XX}(\omega) >> S_{UU}(\omega)$$

对应的那些频率处最小，必须满足 $G \approx H^{-1}$，而在

$$S_{XX}(\omega) << S_{UU}(\omega)$$

所对应的那些频率处，$G \approx 0$。

*9.47　设 $X(t)$为一个平稳高斯随机过程，作为图 P9.47 的系统输入信号。$Z(t)$的功率谱密度由试验测得

$$S_{ZZ}(\omega) = \pi\delta(\omega) + \frac{2\beta}{(\omega^2 + \beta^2)(\omega^2 + 1)}$$

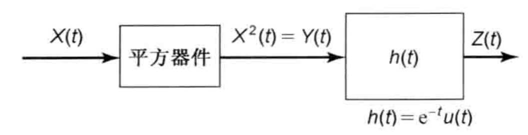


图 P9.47　平方非线性系统后接线性滤波器

(a)求 $Y(t)$的相关函数，用 β 表示。

(b)求 $X(t)$的相关函数。

9.48　考虑一个如图 P9.48 所示的两状态马尔可夫链 $N(t)$，分别取值1和2，当在状态1时，要转移到状态2需要的时间平均率为 $\lambda_1 = 1$，当在状态2时，要转移到状态1需要的时间平均率为 $\lambda_2 = 2$。定义状态概率为 $P_1(t)$ 和 $P_2(t)$，其中 $P_i(t) \triangleq P[N(t) = i]$，$i = 1, 2$。

(a)推导 $P_i(t)$ 的微分方程。

(b)求它们的稳态解。

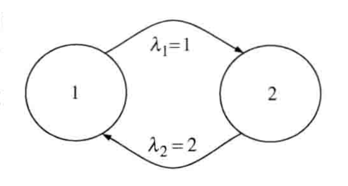


图 P9.48　两状态马尔可夫链状态转移图

9.49　复随机变量的施瓦茨不等式如下：

$$|E[XY^*]| \leqslant \sqrt{E[|X|^2]\ E[|Y|^2]}$$

其中 X 和 Y 为两随机变量。

(a)用施瓦茨不等式来推导联合 WSS 随机过程 $X(t)$ 和 $Y(t)$ 的对应结果:

$$|R_{XY}(\tau)| \leqslant \sqrt{R_{XX}(0)\ R_{YY}(0)}$$

(b)求互功率谱密度的对应结果

$$|S_{XY}(\omega)| \leqslant \sqrt{S_{XX}(\omega)\ S_{YY}(\omega)}$$

提示:根据互功率谱或自功率谱解释(a)的结果。然后引入一个中心频率在任意 ω 处的窄带滤波器。

9.50 维纳过程是对白噪声积分的结果,也被称为布朗运动。设 $B(t)$ 为一个维纳过程,$W(t)$ 为白噪声,可写成

$$B(t) = \int_0^t W(\tau)\mathrm{d}\tau, \qquad t \geqslant 0$$

设 $W(t)$ 为标准白噪声,相关函数为 $R_W(\tau) = \delta(\tau)$。

(a)求互相关函数 $R_{BW}(t_1, t_2)$,并画简图。

(b)求自相关函数 $R_{BB}(t_1, t_2)$,并画简图。

9.51 考虑例 9.2-4 中的两处理器的可靠性问题,一个三状态连续时间马尔可夫随机过程 $X(t)$,其状态转移图参见图 P9.51。$X(t)$ 表示在时刻 t 处理器"工作"的个数。

(a)写出状态概率向量 $\boldsymbol{p}(t)$ 的微分方程

$$\mathrm{d}\boldsymbol{p}(t)/\mathrm{d}t = \boldsymbol{A}\boldsymbol{p}(t)$$

求生成矩阵 $\boldsymbol{A}$ 的确切值。

(b)通过解齐次矩阵-向量方程 $\boldsymbol{Ap} = \boldsymbol{0}$,以及在概率向量 $\boldsymbol{p}$ 中所有概率的和必须为 1 的约束条件下,使得我们可以确定稳态概率向量 $\boldsymbol{p}$。有人认为"概率流"通过图 P9.51 中虚的垂直线一定在稳态时平衡,即 $2\mu p_0 = \lambda p_1$,$\mu p_1 = 2\lambda p_2$,其中 p_i 为向量 $\boldsymbol{p}$ 中的元素,即处在状态 i,$i = 0, 1, 2$ 的稳态概率。请陈述为什么这是一个合理的推断,并通过推导得到的方程满足 $\boldsymbol{Ap} = \boldsymbol{0}$ 来证明它。

(c)求当每小时 $\lambda = 0.001$、$\mu = 0.1$ 时稳态概率的数值解。

9.52 考虑如图 P9.52 所示的三输入、两输出的 LSI 系统。输入为随机过程 $X_1(t)$、$X_2(t)$ 和 $U(t)$,它们是联合广义平稳的并且两两正交,即 $X_1 \perp X_2$, $X_1 \perp U$, $X_2 \perp U$。给定以下函数:系统传递函数 H、G 和 B,以及三输入的功率谱密度 $S_{X_1X_1}$、$S_{X_2X_2}$ 和 S_{UU}。你可将答案用这三个函数来表示。

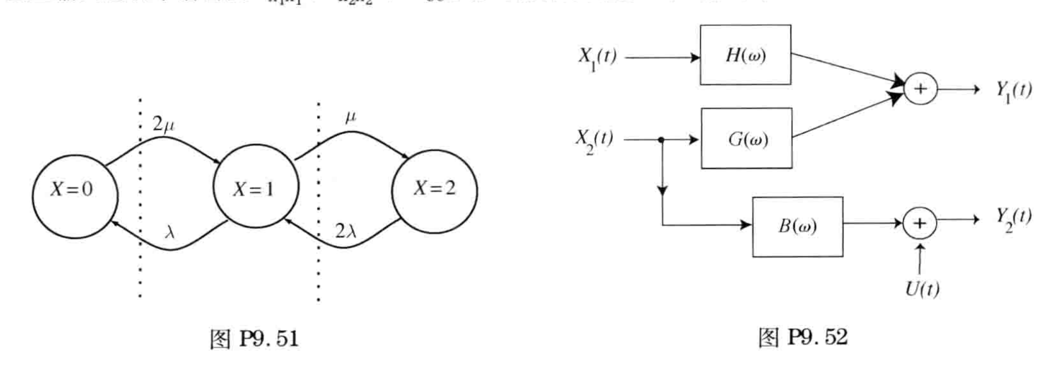


图 P9.51 图 P9.52

(a)求输入/输出互功率谱密度 $S_{Y_1X_1}(\omega)$。

(b)求输入/输出互功率谱密度 $S_{Y_2X_2}(\omega)$。

(c)求输入/输出互功率谱密度 $S_{Y_1Y_2}(\omega)$。

9.53 考虑下面的节拍延迟线问题。假设有一表征节拍的随机序列 A_n 和一个 WSS 随机过程 $X(t)$ 来作为信号模型。假设节拍的总数为 N,节拍的间隔为 T。再假设随机序列 A_n 和随机过程 $X(t)$ 是相互独立的。节拍延迟线的输出为

$$Y(t) = \sum_{n=0}^{N-1} A_n X(t - nT)$$

随机序列 A_n 的相关函数为 $R_A(n_1, n_2)$，WSS 随机过程的相关函数为 $R_X(\tau)$。

(a)求输出相关函数 $R_Y(t_1, t_2)$，用给定函数和参数来表示。

(b)$Y(t)$的广义平稳性是否依赖于随机序列 A_n 也是广义平稳过程？证明你的结论。

(c)在求解(a)的过程中，是否只要给定 A_n 和 $X(t)$不相关的条件就够了，为什么？

9.54　设某无线包信道(吉尔伯特信道模型)有"好"和"坏"两种状态，可用连续时间、两状态的马尔可夫链来建模，其状态转移图如图 P9.54 所示。

(a)求处于"坏"状态的稳态概率。

(b)处于"好"状态时，所有的数据包都被接收。处于"坏"状态时，所有的数据包都被丢失。这会导致丢失一串数据包。在一串数据包丢失时，所有的数据包都丢失。丢失的一串数据包的平均长度为多少？证明它。注意在丢失的一串数据包中，所有数据包都处于"坏"状态。

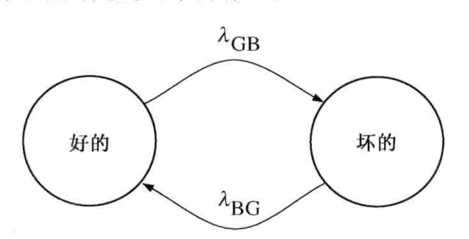


图 P9.54　吉尔伯特信道模型

9.55　考虑一个平均到达率 $\lambda = 3$ 的泊松随机过程 $N(t)$。

(a)求 $N(4) = 2$ 的概率。

(b)求 $N(1) = 1$ 和 $N(2) = 2$ 的联合概率。

9.56　考虑如图 P9.56 所示系统

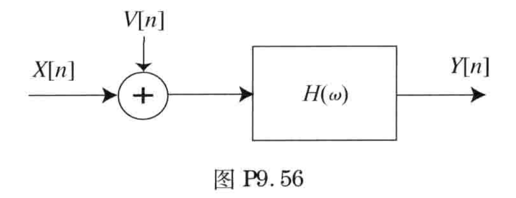


图 P9.56

设 $X[n]$ 和 $V[n]$ 为 WSS，且互不相关，均值都为 0，功率谱密度分别为 $S_{XX}(\omega)$ 和 $S_{VV}(\omega)$。

(a)求输出 $Y[n]$的功率谱密度。

(b)求输入 $X[n]$和输出 $Y[n]$的互谱密度，即 $S_{XY}(\omega)$。

9.57　$X(t)$和 $Y(t)$为两零均值随机过程，已知相关系数为

$$\rho_{XY}(t_1, t_2) \triangleq \frac{E[X(t_1)Y^*(t_2)]}{\sqrt{E[|X(t_1)|^2]E[|Y(t_2)|^2]}}$$

假设平均功率 $E[|X(t)|^2] = E[|Y(t)|^2] \triangleq P$ 为一常数。将增加两随机噪声 $U(t)$和 $V(t)$，它们与 $X(t)$和 $Y(t)$相互正交

$$\widetilde{X}(t) \triangleq X(t) + U(t)$$

$$\widetilde{Y}(t) \triangleq Y(t) + V(t)$$

其中 U 和 V 也相互正交，且均值为 0，平均功率 $E[|U(t)|^2] = E[|V(t)|^2] \triangleq \epsilon$ 为常数。求波浪号过程的相关系数函数，即 $\rho_{\widetilde{X}\widetilde{Y}}(t_1, t_2)$，用原始过程 X 和 Y 的相关系数来表示。

9.58　考虑如图 P9.58 所示系统。两个输入随机序列是 WSS，其各自的功率谱密度为

$$S_{XX}(\omega) = \frac{1}{\omega^2 + 5} \quad \text{和} \quad S_{VV}(\omega) = \frac{2\omega^2 + 8}{(\omega^2 + 3)(\omega^2 + 5)}$$

系统函数 $H(\omega)$ 为 $10[u(\omega + \pi/2) - u(\omega - \pi/2)]$，$-\pi < \omega < +\pi$] 其中 u 为单位阶跃函数，假设 X 和 V 为零均值。

(a)假设 X 和 V 不相关，求输出随机序列 $Y[n]$ 的功率谱密度。

(b)设 X 和 V 的互谱密度为

$$S_{XV}(\omega)=\frac{1}{\omega^2+5}$$

求输出 Y 的功率谱密度。

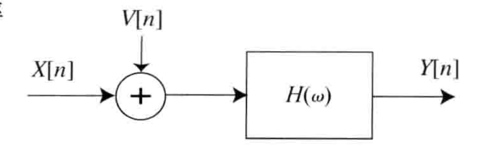


图 P9.58 信号加噪声输入系统

9.59 考虑随机过程 $X(t)=\cos(\omega_0 t+\Theta)$，其中 Θ 为均匀分布在 $[0, 2\pi]$ 上的随机变量，ω_0 为一固定频率。求一阶概率密度 $f_X(x;t)$。此过程是否一阶平稳？求给定 $X(t_1)=x_1$ 时 $X(t_2)$ 的条件概率密度，表示为 $f_X(x_2 \mid x_1;t_1,t_2)$，假设 $t_1<t_2$。

*9.60 设 $Z(t)=X(t)+\mathrm{j}Y(t)$，其中 $X(t)$ 和 $Y(t)$ 为联合广义平稳实随机过程。假设 $X(t)$ 和 $Y(t)$ 为零均值相互正交的过程。定义一个调制在载频 ω_0 上的新随机过程 $U(t)=\mathrm{Re}\{Z(t)\mathrm{e}^{-\mathrm{j}\omega_0 t}\}$。给定各自的相关函数，即 $R_{XX}(\tau)$ 和 $R_{YY}(\tau)$，求使 $U(t)$ 也是 WSS 随机过程的一般条件。说明你的条件有哪些，使产生的过程 $U(t)$ 确实为 WSS。一些有用的三角恒等式如下：

$$\cos(\alpha\pm\beta)=\cos\alpha\cos\beta\mp\sin\alpha\sin\beta$$

$$\sin(\alpha\pm\beta)=\sin\alpha\cos\beta\pm\cos\alpha\sin\beta$$

9.61 求图 P9.61 所示四状态马尔可夫链的稳态概率。将答案用指数率 λ_i 和 μ_i 表示。注意状态是用 1 ~4 来表示的。

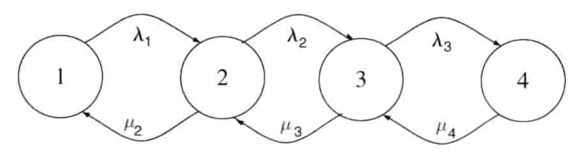


图 P9.61

提示：回忆一下习题 9.51(b)中有关概率流的概念。

9.62 考虑图 P9.62 所示的三状态马尔可夫过程 $X(t)$，状态转移图如图 P9.62 所示。这里状态标签即为真实的输出，即 $X(t)=3$ 时过程处在状态 3。状态转移是由联合独立、呈指数分布的到达时距所控制的，平均速率如图中分支所示。

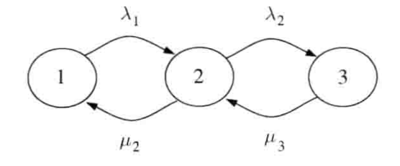


图 P9.62 三状态马尔可夫过程

(a)假设在 $t=0$ 时刻由状态 2 起始，则对任意 $t>0$，在 t 时刻第一次离开这个状态的概率为多少？

(b)求 $t\geqslant 0$ 时关于状态概率的如下向量微分方程的解：

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{p}(t)$$

其中 $\boldsymbol{p}(t)=[p_1,p_2,p_3]^{\mathrm{T}}$，用 λ_i 和 μ_i 表示生成矩阵 $\boldsymbol{A}$。

(c)证明对于 $t\geqslant 0$ 解可表示为

$$\boldsymbol{p}(t)=\exp(\boldsymbol{A}t)\,\boldsymbol{p}(0)$$

其中 $\boldsymbol{p}(0)$ 为初始概率向量，矩阵 $\exp(\boldsymbol{A}t)$ 由无穷级数定义为

$$\exp(\boldsymbol{A}t)\triangleq\boldsymbol{I}+\boldsymbol{A}t+\frac{1}{2!}(\boldsymbol{A}t)^2+\frac{1}{3!}(\boldsymbol{A}t)^3+\frac{1}{4!}(\boldsymbol{A}t)^4+\cdots$$

不必担心此级数的收敛性，但已知其对所有有限的 t 值，它是绝对收敛的。

参考文献

9-1 S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes.* New York: Academic Press, 1975.

9-2 L. Kleinrock, *Queueing Systems*, Vol. 1: *Theory.* New York: John Wiley, 1975.

9-3 T. Kailath, *Linear Systems.* Upper Saddle River, NJ: Prentice Hall, 1980.

9-4 E. W. Kamen, *Introduction to Signals and Systems*, 2nd edition. New York: Macmillan, 1990, p. 172.

9-5 E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems.* New York: Springer-Verlag, 1985, pp. 62-63.

9-6 P. Billingsley, *Probability and Measure*, New York: John Wiley, 1979, pp. 467-477.

9-7 J. L. Doob, *Stochastic Processes.* New York: John Wiley, 1953.

9-8 P. Whittle, "On Stationary Process in the Plane," *Biometrika*, Vol. 41 (1954), pp. 434-449.

9-9 A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd Edition. Upper Saddle River, NJ: Prentice Hall, 1999, Chapters 2-3.

附录 A　相关数学知识复习

本附录将复习一些数学知识，这些知识可用于概率论和随机过程的学习。我们先从离散数学和连续数学的基本概念开始。

A.1　数学基础

先复习序列的基本概念并给出一些例子，随后再复习序列求和，最后复习 Z 变换。

序列

序列就是一组整数集往实数(或复数)集的映射。一般来说，整数集由非负整数($n \geqslant 0$)或所有整数($-\infty < n < \infty$)组成。

一个经常遇到的例子就是指数序列 a^n ($n \geqslant 0$)，图 A.1-1 给出了实数 a 取不同值时该序列的图形。注意到当 $|a| > 1$ 时该序列是发散的，当 $|a| < 1$ 时，该序列收敛到 0。当 $a = 1$ 时，该序列是常数 1，当 $a = -1$ 时，该序列交替取值 +1 和 -1。

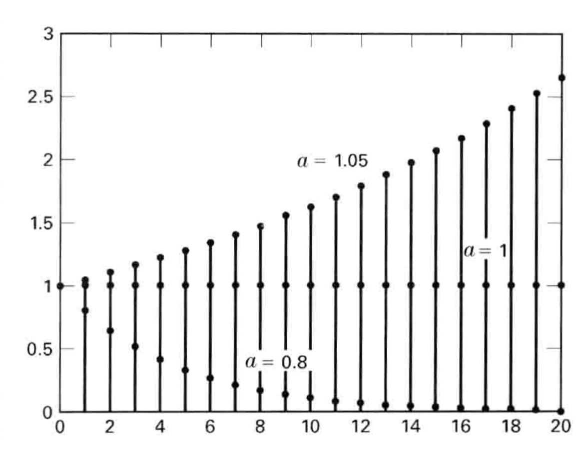


图 A.1-1　a 分别取 1.05，1.0 和 0.8 时指数序列图形

与之相关的一个重要序列就是复指数序列 $\exp(\mathrm{j}\omega n)$，该序列是线性时不变系统的特征函数。这意味着对于一个频率响应函数为 $H(\omega)$ 的系统，如果输入是 $\exp(\mathrm{j}\omega n)$，那么对应的输出就是 $H(\omega)\exp(\mathrm{j}\omega n)$，正好对输入进行了缩放。

收敛性

一个定义在正整数($n \geqslant 1$)上的序列(用 $x[n]$ 或 x_n 来表示)，如果随着 n 的增大 $x[n]$ 的取值逐渐趋向于一个有限值 x，那么该序列就收敛到 x。更准确的定义是：任取一个 $\varepsilon > 0$，一定存在一个值 $N_0(\varepsilon)$，当 $n \geqslant N_0$ 时，$|x[n] - x| < \varepsilon$。需要注意的是，N_0 的取值取决于 ε。

例 A.1-1　序列 a_n 的定义如下：

$$a_n = 2^n/(2^n + 3^n)$$

试找出 $n \to \infty$ 时序列的极限。通过观察可知，该序列的极限为 $a_n = 0$。为了完成证明，通过上式可以将 $N_0(\varepsilon)$ 的表达式写为

$$\left|\frac{2^n}{2^n + 3^n}\right| < \varepsilon$$

或者有

$$N_0(\varepsilon)=\frac{\ln\left(\frac{1-\varepsilon}{\varepsilon}\right)}{\ln\frac{3}{2}}$$

在这里我们假定 $0<\varepsilon<1$。对于 ε 任意一个确定的取值，N_0 的取值是一个有限值。

求和

在我们的工作中经常会遇到对序列求和。几何序列 a^n 经常被用来讲解如何对序列求和。容易得到下面这个求和公式，在此假设 $n_2\geqslant n_1$。

$$\sum_{n=n_1}^{n_2}a^n=\frac{a^{n_1}-a^{n_2+1}}{1-a},\qquad a\neq 1 \tag{A.1-1}$$

很显然，当 $a=1$ 时，该序列和正好为 n_2-n_1+1。有一个简单方法可以用来验证式(A.1-1)的正确性，首先定义 $S=\sum_{n_1}^{n_2}a^n$，然后再利用几何序列的特殊性质得到

$$aS=S+a^{n_2+1}-a^{n_1}$$

此时，通过解上面这个方程就可以得到 S 的表达式 $(a\neq 1)$，即式(A.1-1)。

当 $|a|\leqslant 1$ 时，该求和公式的上限 n_2 可以趋向于 ∞，此时得到

$$\sum_{n=n_1}^{\infty}a^n=\frac{a^{n_1}}{1-a},\qquad |a|<1 \tag{A.1-2}$$

另外一个有用的求和公式如下：

$$\sum_{n=n_1}^{\infty}na^n=\frac{n_1a^{n_1}(1-a)+a^{n_1+1}}{(1-a)^2},\qquad |a|<1 \tag{A.1-3}$$

对于式(A.1-2)和式(A.1-3)，求和公式的下限经常是 $n_1=0$。

Z 变换

该变换对于分析线性时不变系统的特性，以及求解线性恒系数差分方程非常有用。确定序列 $x[n]$ 的 Z 变换定义如下：

$$\boldsymbol{X}(z)=\sum_{n=-\infty}^{+\infty}x[n]z^{-n},\qquad z\in\mathscr{R}$$

在上面这个方程中，区间 $\mathscr{R}$ 被称为收敛区间，该区间代表复数 z 的取值范围，同时也表示 Z 变换的定义域。在这个区间 $\mathscr{R}$ 内，对应的 z 和 $x[n]$ 的和绝对收敛，即

$$\sum_{n=-\infty}^{+\infty}|x[n]||z|^{-n}<\infty$$

区间 $\mathscr{R}$ 的一般表达式为 $\mathscr{R}=\{z:R_-<|z|<R_+\}$，是一个环状区间。集合 $\{z|R_-<|z|<R_+\}$ 的可被理解为“该集合中所有点 z 的幅度(长度)都大于 R_- 小于 R_+”。

例 A.1-2　离散时间序列 $x[n]$ 为指数形式，定义如下：

$$x[n]=a^n\exp(\mathrm{j}\omega_0n)u[n]$$

$u[n]$ 表示单位阶跃序列，当 $n\geqslant0$ 时 $u[n]=1$，当 $n<0$ 时 $u[n]=0$。

试计算该序列的 Z 变换，步骤如下：

$$\begin{aligned}\boldsymbol{X}(z) &= \sum_{n=0}^{\infty} a^n \exp(\mathrm{j}\omega_0 n) z^{-n} \\ &= \sum_{n=0}^{\infty} (a\mathrm{e}^{\mathrm{j}\omega_0} z^{-1})^n \\ &= \frac{1}{1 - a\mathrm{e}^{\mathrm{j}\omega_0} z^{-1}}, \qquad |z| > |a|, \quad \mathscr{R} = \{z : |a| < |z|\}\end{aligned} \tag{A.1-4}$$

Z 变换在离散时间信号处理中作用非常大，这是因为 Z 变换中关于卷积和乘法的基本原理。

定理 A.1-1 两个绝对可和的序列 $x[n]$ 和 $h[n]$ 相卷积，可产生一个新的序列 $y[n]$ 如下：

$$y[n] = \sum_{m=-\infty}^{+\infty} x[m]h[n-m]$$

在此我们用 $y = h * x$ 来表示这种操作。那么 $y[n]$ 的 Z 变换就可以由 x 和 h 的 Z 变换给出

$$\boldsymbol{Y}(z) = \boldsymbol{H}(z)\boldsymbol{X}(z), \qquad z \in \mathscr{R}_h \cap \mathscr{R}_x$$

由于序列 h 和 x 都是绝对可和的，那么它们各自的收敛区间 $\mathscr{R}_h$ 和 $\mathscr{R}_x$ 都应该包含 z 平面上的单位圆(即 $\{|z| = 1\}$)。那么 Z 变换 $\boldsymbol{Y}(z)$ 的定义域就为 $z \in \mathscr{R}_h \cap \mathscr{R}_x$，此时应确保该定义域非空。

利用这个结论得到卷积的 Z 变换后，又可以通过 Z 逆变换来得到输出序列 $y[n]$。有许多方法可以求 Z 逆变换，包括把 $\boldsymbol{Y}(z)$ 展开成幂级数的形式，如果 $\boldsymbol{Y}(z)$ 是一个 z 的多项式相除的形式还可以用长除法，以及最有效的留数法。在本附录 A.3 节还会介绍如何使用留数法来求拉普拉斯逆变换。

A.2 连续数学

本节主要复习有关一重实数和二重实数积分的计算。

定积分和不定积分

在微积分课程中，我们学习了定积分和不定积分两种积分，定义如下：

$$\int x^2 \mathrm{d}x = \frac{1}{3}x^3 + C \qquad \text{不定积分}$$

$$\int_a^b x^2 \mathrm{d}x = \frac{1}{3}b^3 - \frac{1}{3}a^3 \qquad \text{定积分}$$

在本节中基本上只会提及定积分，而很少涉及不定积分。这是因为我们需要用积分来衡量特定某种特性，而不仅仅通过给定的导数来求原函数。请注意这两种积分的区别。不同于不定积分，定积分是其积分上限和下限的函数，不是 x 的函数。因此有的时候我们把定积分中的 x 看成“形式变量”，这是因为 x 完全可以用其他变量(如 y)来取代，而这完全不会影响定积分的结果，也就是

$$\int_a^b x^3 \mathrm{d}x = \int_a^b y^3 \mathrm{d}y$$

为了计算定积分的结果，一般先求出不定积分的原函数，然后把原函数在积分上限的值减去积分下限值即可。

在普通微积分课程中，一般不会强调定积分对集合的操作，甚至有与“曲线下方面积”无关的定积分存在，这就是所谓的黎曼积分(Riemann)。以下面的定积分为例：

$$I = \int_a^b f(x)\mathrm{d}z(x)$$

这里 x 的定义域集合为 $\{x: a \leqslant x \leqslant b\}$，通过将区间$(a, b)$划分为 n 等份得到 $\Delta x \equiv (b-a)/n$，然后再计算每份数值结果来求整个区间的积分值，也即是

$$I_n(a,b) = \sum_i f(i\Delta x) \times \Big(z(i\Delta x + \Delta x/2) - z(i\Delta x - \Delta x/2)\Big) \xrightarrow[\substack{\Delta x \to 0, \\ n \to \infty}]{} I$$

其中 $i\Delta x$，$i\Delta x \pm \Delta x/2 \in \{x: a \leqslant x \leqslant b\}$。如果 $z(x) = x$，那么此时 I 就变为了非常著名的“曲线下方面积”的黎曼积分。但在有些情况下，黎曼积分的条件并不满足。比如第 4 章中求数学期望的操作 ($E[x] = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x$)，如果结果能够收敛到一个有限值，那么就要求函数 $f_X(x)$ 只存在有限个断点。如果 $f_X(x)$ 不满足这种要求，我们仍然可以通过 $E[X] = \int_{-\infty}^{\infty} x\mathrm{d}F_X(x)$ 来求 $E[X]$，其中 $F_X(x)$ 表示 X 的累积分布函数(CDF)。这种类型的积分称为 Stieljes 积分，这也是“曲线下方面积”类型的积分，这部分知识往往会在微积分课程开头讲到。例如，如果 $F_X(x) = (1 - \mathrm{e}^{-\lambda x})u(x)$，那么 $\mathrm{d}F_X(x) = (\lambda \mathrm{e}^{-\lambda x})u(x)\mathrm{d}x$，可以得到 $E[X] = \int_0^{\infty} x\lambda \mathrm{e}^{-\lambda x}\mathrm{d}x = 1/\lambda$。

积分的导数

有的时候需要对积分关于某个变量求导，这个变量可以是积分上限，积分下限，或者积分本身

$$\frac{\mathrm{d}}{\mathrm{d}y}\int_{a(y)}^{b(y)} f(x,y)\mathrm{d}x = f(b(y),y)\frac{\mathrm{d}b(y)}{\mathrm{d}y} - f(a(y),y)\frac{\mathrm{d}a(y)}{\mathrm{d}y} + \int_{a(y)}^{b(y)} \frac{\partial f(x,y)}{\partial y}\mathrm{d}x \quad \text{(A. 2-1)}$$

上面这个重要公式的推导需要用到方程 $I(a, b, y)$，可以把 $b = b(y)$ 和 $a = a(y)$ 都看成 y 的函数，那么有

$$\frac{\mathrm{d}I}{\mathrm{d}y} = \frac{\partial I}{\partial b}\frac{\mathrm{d}b}{\mathrm{d}y} + \frac{\partial I}{\partial a}\frac{\mathrm{d}a}{\mathrm{d}y} + \frac{\partial I}{\partial y}$$

用 I 来代表下面的积分结果：

$$I \triangleq \int_{a(y)}^{b(y)} f(x,y)\mathrm{d}x$$

以及定义一个函数 $F(x, y)$，可得

$$f(x,y) \triangleq \frac{\partial F(x,y)}{\partial x}$$

得到如下结果：

$$\frac{\partial I}{\partial b}=f(b(y),y)$$
$$\frac{\partial I}{\partial a}=-f(a(y),y)$$
$$\frac{\partial I}{\partial y}=\frac{\partial}{\partial y}\int_{a(y)}^{b(y)}f(x,y)\mathrm{d}x=\int_{a(y)}^{b(y)}\frac{\partial}{\partial y}f(x,y)\mathrm{d}x$$

上面最后一项等号的右边应该把 $b(y)$ 和 $a(y)$ 当成常数，这是因为前面两项在计算 I 的变化范围时已经把积分上下限计算出来了。

下面给出一个例题来讲解如何使用上面的公式，这个例题在概率函数通过系统变换的学习中出现过。

例 A.2-1 试考虑函数 $f(x, y) = x + 2y$ 的变换

$$\frac{\partial}{\partial y}\int_0^y (x+2y)^2\mathrm{d}x=(y+2y)^2 1-(0+2y)^2 0+\int_0^y 4(x+2y)\mathrm{d}x$$
$$=(3y)^2+4\left(\frac{1}{2}x^2+2yx\right)\Big|_0^y$$
$$=(3y)^2+2y^2+8y^2=19y^2$$

分部积分法

分部积分法是积分计算中一个非常有用的方法，公式如下：

$$\int_a^b u(x)\mathrm{d}v(x)=u(x)v(x)|_a^b-\int_a^b v(x)\mathrm{d}u(x) \tag{A.2-2}$$

其中 u 和 v 表示积分中关于 x 的函数，x 的取值范围为 $a \leqslant x \leqslant b$。该公式的推导利用了求导的链条法则，即对函数乘积 $u(x)v(x)$ 的求导。下面会给出一个例子来讲解。通过分部积分法，可以求出很多不容易求解的积分解析表达式。

例 A.2-2 试求解下面的积分：

$$\int_0^\infty x\mathrm{e}^{-2x}\mathrm{d}x$$

令 $u(x) = x$, $\mathrm{d}v(x) = \mathrm{e}^{-2x}\mathrm{d}x$, 然后利用分部积分公式可得

$$\int_0^\infty x\mathrm{e}^{-2x}\mathrm{d}x=x\left(-\frac{1}{2}\mathrm{e}^{-2x}\right)\Big|_0^\infty-\int_0^\infty\left(-\frac{1}{2}\mathrm{e}^{-2x}\right)\mathrm{d}x$$
$$=\frac{1}{2}\int_0^\infty \mathrm{e}^{-2x}\mathrm{d}x=\frac{1}{2}\left(-\frac{1}{2}\mathrm{e}^{-2x}\right)\Big|_0^\infty$$
$$=\frac{1}{4}$$

完全平方法

通过完全平方法，可以利用已知的积分值来求未知的积分值。比如，要凑出公式 x^2+4x 的完全平方表达式，可以将该公式加 4 得到 $(x+2)^2$，然后再减去 4，即

$$x^2+4x=(x+2)^2-4$$

下面来演示如何用该方法来求积分值。在此以本书中经常遇到的高斯积分为例

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}x^2}\mathrm{d}x = \sqrt{2\pi}$$

假如需要求如下的积分：

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}(x^2+4x)}\mathrm{d}x = ?$$

此时就可以用完全平方法如下：

$$\mathrm{e}^2\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}(x^2+4x+4)}\mathrm{d}x$$

在这里，我们在积分中乘以了 e^{-2}，在积分外又乘了一个 e^2，然后继续

$$= \mathrm{e}^2\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}(x+2)^2}\mathrm{d}x$$

进行变量替换 $y = x + 2$，等式变为

$$\begin{aligned} &= \mathrm{e}^2\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}y^2}\mathrm{d}y \\ &= \mathrm{e}^2\sqrt{2\pi} \end{aligned}$$

二重积分

在(x, y)平面上的积分就称为二重积分。这里的积分微元是一个面积 $\mathrm{d}x\mathrm{d}y$。这种积分的求解应该按照一定的顺序，比如先 x 再 y，或者先 y 再 x，因此这种积分是重积分。有三种可能的表达式如下：

$$\int_{y_1}^{y_2}\left(\int_{x_1}^{x_2} f(x,y)\mathrm{d}x\right)\mathrm{d}y = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f(x,y)\mathrm{d}x\,\mathrm{d}y = \int_{x_1}^{x_2}\left(\int_{y_1}^{y_2} f(x,y)\mathrm{d}y\right)\mathrm{d}x$$

中间的积分是二重积分或者面积元积分。由于在积分运算中应该限制运算量，因此有一个问题存在：上面的三个二重积分一直都是等价的吗？幸运的是，在测量理论[9-1]中有一个非常有用的结论，那就是：如果一个积分满足 Lebesgue 可积函数的定义，那么上面三种积分表达式都是等价的，或者三者都不存在。在此我们仅考虑最常见的情况，即三者表达式都存在且等价的时候。

需要注意的是，在等式最左边是对 x 先积分，而等式最右边是对 y 先积分。而对于中间的这个二重积分采用的标记读为 $x-y$，这与二元函数的输入参数或面积微元 $\mathrm{d}x\mathrm{d}y$ 相类似，这样就不会与下面的二重积分相混淆。

$$\int_1^3\int_0^5 x\mathrm{e}^{-y}\mathrm{d}x\,\mathrm{d}y$$

因为我们应该知道，这个积分区域是以点 $(x, y) = (1, 0)$ 和点 $(x, y) = (3, 5)$ 为对角的矩形区域。

函数

函数是指从定义域 $\mathscr{X}$ 往值域 $\mathscr{Y}$ 的单一映射。函数定义中要求的“单一”，是指每个 x 只能得到一个 y，也即是$f(x)$只能有且仅有一个结果。$f(x) = x^2$ 就是一个函数，而$f(x) = \pm\sqrt{x}$ 就不是一个函数。

单调函数　关于实变量 x 的单调函数是指函数值随着 x 的增大而一直增大，或者随着 x 的增大而一直减小的函数。前者(斜率为正)称为单调递增，后者(斜率为负)称为单调递减，如图 A.2-1 和图 A.2-2 所示。如果一个单调函数在某些区域的斜率为 0，那么这种函数就是单调不减或单调不增的，如图 A.2-3 所示。

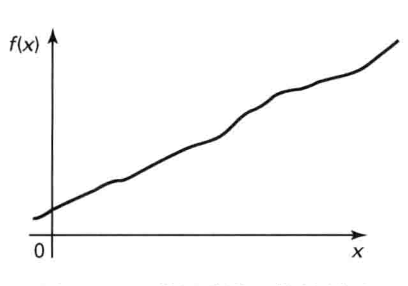


图 A.2-1　单调递增函数的例子

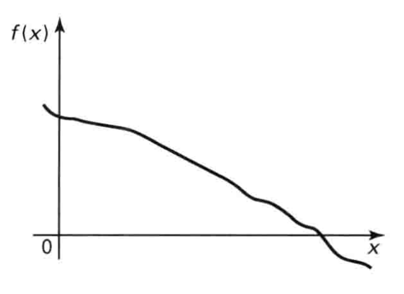


图 A.2-2　单调递减函数的例子

逆函数

一个函数的逆函数可能存在也可能不存在。一个函数存在逆函数，还要求原函数为一一映射，即值域 $\mathscr{Y}$ 中任意一个 y 在定义域 $\mathscr{X}$ 中只有唯一的 x 与之对应。这样我们才能定义一个从 $\mathscr{Y}$ 映射到 $\mathscr{X}$ 的逆函数 $f^{-1}(y)$。同时可知，逆函数存在的充要条件就是原函数 $f(x)$ 是单调递增或单调递减的。图 A.2-3 中的函数就没有逆函数，这是因为该函数存在一截斜率为 0 的平坦区域。

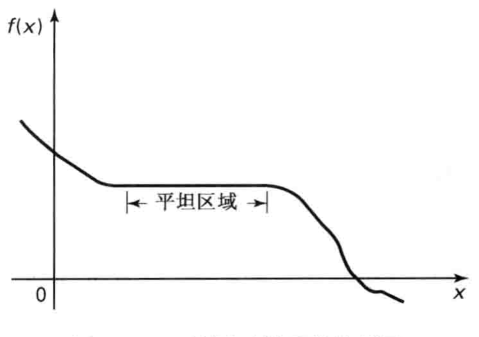


图 A.2-3　单调不增函数的例子

A.3　留数法求傅里叶逆变换①

在第 8 章和第 9 章，我们分别定义了离散时间序列和连续时间序列的功率谱(PSD) $S(\omega)$，同时也表明功率谱对于分析输入为随机过程的 LTI 非常有用。我们常常需要对给定的功率谱求傅里叶逆变换来得到它的相关函数，这样就可以分析它的时域特性。本节将复习如何基于留数法求傅里叶逆变换。

我们先复习广义平稳(WSS)随机序列的功率谱密度和相关函数之间的变换关系

$$S(\omega)=\int_{-\infty}^{+\infty} R(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau$$

$$R(\tau)=\frac{1}{2\pi}\int_{-\infty}^{+\infty} S(\omega)\mathrm{e}^{+\mathrm{j}\omega\tau}\,\mathrm{d}\tau$$

为了求上述公式的傅里叶逆变换，需要采用基于复变量理论[A-3]的留数法，那么首先必须把该积分表示成一个沿 s 平面的围线积分形式。在此我们定义一个关于复变量 $s=\sigma+\mathrm{j}\omega$ 的新函数 $\boldsymbol{S}$。

① 假设阅读本节内容的读者非常熟悉第 8 章和第 9 章的内容。

首先，在虚轴上定义 $\boldsymbol{S}(s)$，这样就可以表示成实变量方程 $S(\omega)$ 的形式

$$\boldsymbol{S}(s)|_{s=\mathrm{j}\omega} \triangleq S(\omega)$$

然后，用变量 s 来取代 $\mathrm{j}\omega$，将函数 $\boldsymbol{S}(\mathrm{j}\omega)$ 扩展到整个复平面。

$$\boldsymbol{S}(s)|_{s=\mathrm{j}\omega} = S(\omega) = \int_{-\infty}^{+\infty} R(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\,\mathrm{d}\tau$$

因此有

$$\boldsymbol{S}(s) = \int_{-\infty}^{+\infty} R(\tau)\mathrm{e}^{-s\tau}\,\mathrm{d}\tau \tag{A.3-1}$$

上式为相关函数 R 的双边拉普拉斯变换。同样地，由傅里叶逆变换

$$\begin{aligned} R(\tau) &= \frac{1}{2\pi\mathrm{j}}\int_{-\infty}^{+\infty} \boldsymbol{S}(s)|_{s=\mathrm{j}\omega}\mathrm{e}^{\mathrm{j}\omega\tau}\mathrm{d}(\mathrm{j}\omega) \\ &= \frac{1}{2\pi\mathrm{j}}\int_{-\mathrm{j}\infty}^{+\mathrm{j}\infty} \boldsymbol{S}(s)\mathrm{e}^{s\tau}\,\mathrm{d}s \end{aligned} \tag{A.3-2}$$

上式是 s 平面上沿虚轴的积分。

根据复变函数理论[A-2][A-3]，式(A.3-2)中的积分称为围线积分，利用留数法可以计算这种封闭围线的积分值。当被积函数是有理函数时，这种方法非常简单，其中有理函数是指该函数可以表示成 s 的两个多项式相除的形式。由于在以差分方程描述的线性系统中这个条件很容易满足，因此用留数法求积分值的方法非常有用。我们从复变量理论出发来给出主要的结论。

事实

假设 $\boldsymbol{F}(s)$ 是复变量 s 的函数，$\boldsymbol{F}(s)$ 在 C 逆时针方向包围的区域内解析，且在该区域内存在 P 个孤立的奇点。这些极点分别为 $s = p_i$，$i = 1, \cdots, P$。那么有

$$\frac{1}{2\pi\mathrm{j}}\oint_C \boldsymbol{F}(s)\mathrm{d}s = \sum_{\substack{p_i \ \text{inside} \\ C}} \mathrm{Res}[\boldsymbol{F}(s); s = p_i] \tag{A.3-3}$$

其中

1. 对于一阶极点，$\mathrm{Res}[\boldsymbol{F}(s); s = p] = [\boldsymbol{F}(s)(s-p)]|_{s=p}$
2. 对于二阶极点，$\mathrm{Res}[\boldsymbol{F}(s); s = p] = \dfrac{\mathrm{d}}{\mathrm{d}s}[\boldsymbol{F}(s)(s-p)^2]|_{s=p}$
3. 对于 n 阶极点，$\mathrm{Res}[\boldsymbol{F}(s); s = p] = \dfrac{1}{(n-1)!}\left(\dfrac{\mathrm{d}^{(n-1)}}{\mathrm{d}s^{(n-1)}}[\boldsymbol{F}(s)(s-p)^n]\right)\Big|_{s=p}$

为了应用上面的结论来求解问题，我们需要选取合适的积分路线 C_L。如果选取的积分路线是半圆(如图 A.3-1 所示)，那么当 $|s|\to+\infty$ 时函数 $\boldsymbol{S}(s)\mathrm{e}^{s\tau}$ 将会很快趋向于0，那么在该区域内函数 $S(s)\mathrm{e}^{s\tau}$ 的积分值不会变化。换句话说，函数 $\boldsymbol{S}(s)\mathrm{e}^{s\tau}$ 在这个半圆区域内的积分值为0。该结论成立的条件是当 $|s|\to+\infty$ 时，函数 $|\boldsymbol{S}(s)|$ 有界，即

$$|\mathrm{e}^{s\tau}| \to 0 \quad 当 \quad \mathrm{Re}(s) \to -\infty$$

其中上式后半部分成立的前提是 $\tau > 0$。因此对于所有的正值 τ，可以得到

$$R(\tau)=\frac{1}{2\pi j}\oint_{C_L}\boldsymbol{S}(s)\mathrm{e}^{s\tau}\mathrm{d}s=\sum_{\substack{p_i\\ \text{inside } C_L}}\mathrm{Res}[\boldsymbol{S}(s)\mathrm{e}^{s\tau};s=p_i]$$

类似地，当 $\tau<0$ 时，应该选取的积分路线是 s 平面的右半部分(如图 A.3-2 所示)，在这种情况下可以得到

$$|\mathrm{e}^{s\tau}|\to 0 \quad \text{当} \quad \mathrm{Re}(s)\to+\infty$$

所以下式成立：

$$R(\tau)=\frac{1}{2\pi \mathrm{j}}\oint_{C_R}\boldsymbol{S}(s)\mathrm{e}^{s\tau}\mathrm{d}s=-\sum_{\substack{p_i\\ \text{inside } C_R}}\mathrm{Res}[\boldsymbol{S}(s)\mathrm{e}^{s\tau};s=p_i]$$

此时，顺时针的曲线意味着 $\tau<0$。

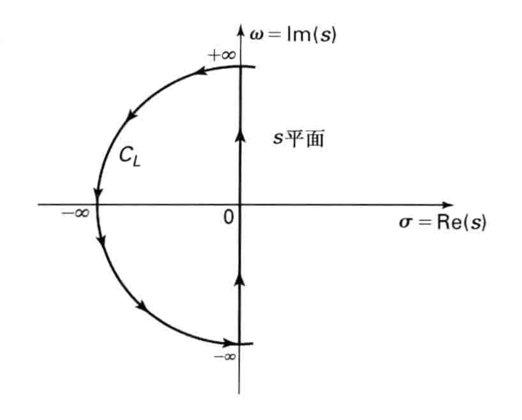


图 A.3-1 s 平面上左半部分闭曲线，$\tau>0$

图 A.3-2 s 平面上右半部分闭曲线，$\tau<0$

例 A.3-1 (一阶功率谱)令

$$S(\omega)=2\alpha/(\alpha^2+\omega^2),\qquad 0<\alpha<1$$

那么有

$$\boldsymbol{S}(s)|_{s=\mathrm{j}\omega}=S(\omega)=2\alpha/(\alpha^2+\omega^2)=2\alpha/(\mathrm{j}\omega+\alpha)(-\mathrm{j}\omega+\alpha)$$

即

$$\boldsymbol{S}(s)=\frac{2\alpha}{(s+\alpha)(-s+\alpha)}$$

上式在 s 平面上的零极点分布情况如图 A.3-3 所示。

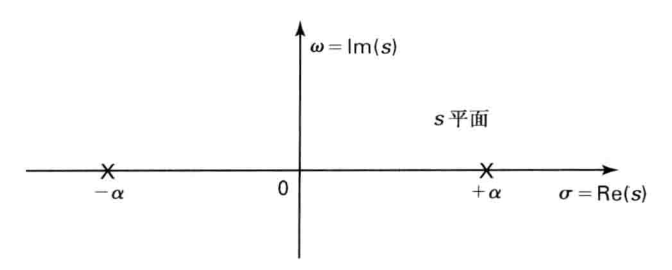


图 A.3-3 零极点图

为了计算 $\tau>0$ 时的留数值

$$R(\tau)=\mathrm{Res}[\boldsymbol{S}(s)\mathrm{e}^{s\tau};s=-\alpha]=\left.\frac{2\alpha\mathrm{e}^{s\tau}}{(-s+\alpha)}\right|_{s=-\alpha}$$
$$=\frac{2\alpha}{2\alpha}\mathrm{e}^{-\alpha\tau}$$

当 $\tau<0$ 时

$$R(\tau)=-\mathrm{Res}[\boldsymbol{S}(s)\mathrm{e}^{s\tau};s=+\alpha]$$
$$=-\left.\frac{2\alpha\mathrm{e}^{s\tau}(s-\alpha)}{(s+\alpha)(-s+\alpha)}\right|_{s=+\alpha}$$
$$=\left.\frac{-2\alpha\mathrm{e}^{s\tau}}{(s+\alpha)(-1)}\right|_{s=\alpha}=\frac{2\alpha}{2\alpha}\mathrm{e}^{\alpha\tau}$$

将上面两种情况的结论合并为简单的表达式，得到

$$R(\tau)=\exp(-\alpha|\tau|),\qquad -\infty<\tau<+\infty$$

随机过程功率谱密度的的傅里叶逆变换

对于一个随机过程，也可以在 z 平面上对其进行围线积分。在此我们先复习一个序列的傅里叶变换及其逆变换公式。

$$S(\boldsymbol{\omega})=\sum_{m=-\infty}^{+\infty}R[m]\mathrm{e}^{-\mathrm{j}\omega m}$$
$$R[m]=\frac{1}{2\pi}\int_{-\pi}^{+\pi}S(\boldsymbol{\omega})\mathrm{e}^{+\mathrm{j}\omega m}\mathrm{d}\boldsymbol{\omega}$$

上式中第二个积分表达式也可以写成复平面上沿单位圆的围线积分，只需要定义一个复变量的函数表达式 $\boldsymbol{S}(z)|_{z=\mathrm{e}^{\mathrm{j}\omega}}\triangleq S(\boldsymbol{\omega})$，用 z 来替换 $\mathrm{e}^{\mathrm{j}\omega}$ 得到一个新的表达式，同时也得到 Z 变换形式的功率谱密度，具体表达式如下：

$$\boldsymbol{S}(z)=\sum_{m=-\infty}^{+\infty}R[m]z^{-m}$$
$$R[m]=\frac{1}{2\pi\mathrm{j}}\oint_C\boldsymbol{S}(z)z^{m-1}\mathrm{d}z\qquad 其中\ C=\{|z|=1\}\tag{A.3-4}$$

在上面这种情况下，积分沿着逆时针方向围绕原点形成闭合曲线，因此我们可以直接应用式(A.3-3)的结论得到

$$R[m]=\sum_{\substack{p_i\\ \text{inside }C}}\mathrm{Res}[\boldsymbol{S}(z)z^{m-1};z=p_i]$$

上式需对单位圆内的所有的极点求和。这个公式对所有的整数 m 有效，不过当 m 为负数的时候这个公式就不好用了，因为当 $z=0$ 的时候极点 z^{m-1} 的值为无穷大。幸运的是，用 z 到 $1/z$ 的一个简单映射就可以很方便地解决这个问题，因此可以得到[A-1]

$$R[m]=\frac{1}{2\pi\mathrm{j}}\oint_C\boldsymbol{S}(z^{-1})z^{-m+1}(-z^{-2}\mathrm{d}z)$$
$$=\frac{1}{2\pi\mathrm{j}}\oint_C\boldsymbol{S}(z^{-1})z^{-m-1}\mathrm{d}z$$

这样就可以回避当 $m<0$ 时极点无穷大的麻烦。综上所述，可以得到如下结论：

当 $m \geqslant 0$ 时

$$R[m] = \sum_{i:\text{单位圆内的极点}} \text{Res}[\boldsymbol{S}(z)z^{m-1}; z = p_i]$$

当 $m < 0$ 时

$$R[m] = \sum_{i:\text{单位圆内的极点}} \text{Res}[\boldsymbol{S}(z^{-1})z^{-m-1}; z = p_i^{-1}]$$

例 A.3-2 (随机序列的一阶功率谱密度)给定一个功率谱密度函数如下:

$$S(\boldsymbol{\omega}) = \frac{2(1-\rho^2)}{(1+\rho^2) - 2\rho\cos\boldsymbol{\omega}}, \qquad |\boldsymbol{\omega}| \leqslant \pi \tag{A.3-5}$$

该函数如图 A.3-4 所示。

利用式 $\cos\omega = 0.5(\mathrm{e}^{\mathrm{j}\omega} + \mathrm{e}^{-\mathrm{j}\omega})$,可将式(A.3-5)转换为复变函数表达式

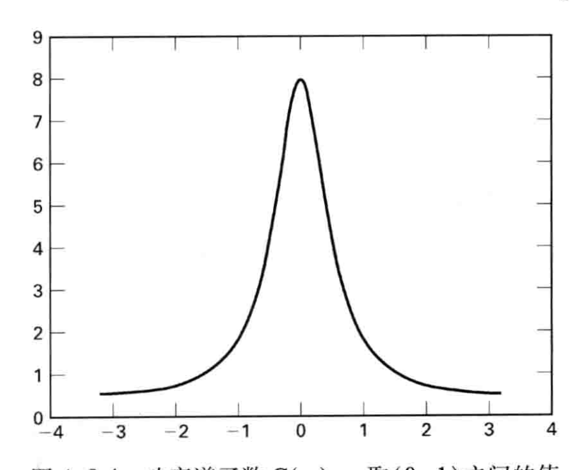
$$\begin{aligned}\boldsymbol{S}(z)|_{z=\mathrm{e}^{\mathrm{j}\omega}} = S(\boldsymbol{\omega}) &= \frac{2(1-\rho^2)}{(1+\rho^2) - 2\rho\cos\boldsymbol{\omega}} \\ &= \frac{2(1-\rho^2)}{(1+\rho^2) - \rho(\mathrm{e}^{+\mathrm{j}\omega} + \mathrm{e}^{-\mathrm{j}\omega})}\end{aligned}$$

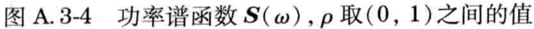
用 z 来取代 $\mathrm{e}^{\mathrm{j}\omega}$ 得到 z 的函数

$$\begin{aligned}\boldsymbol{S}(z) &= \frac{2(1-\rho^2)}{(1+\rho^2) - \rho(z+z^{-1})} \\ &= -2(\rho^{-1}-\rho)\frac{z}{(z-\rho)(z-\rho^{-1})}\end{aligned}$$

上式在 z 平面上的零极点分布如图 A.3-5 所示。在此给出从 $\boldsymbol{S}(\omega)$ 到 $\boldsymbol{S}(z)$ 的转换公式

$$\cos\boldsymbol{\omega} \leftarrow \tfrac{1}{2}(z+z^{-1}) \tag{A.3-6}$$




图 A.3-4 功率谱函数 $\boldsymbol{S}(\omega)$,ρ 取(0, 1)之间的值

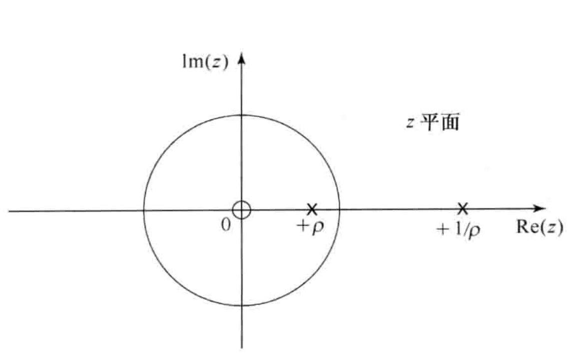


图 A.3-5 z 平面

当 $m \geqslant 0$ 时

$$\begin{aligned}R[m] &= \text{Res}[\boldsymbol{S}(z)z^{m-1}; z = \rho] \\ &= -2(\rho^{-1}-\rho)\frac{\rho\rho^{m-1}}{(\rho-\rho^{-1})} \\ &= 2\rho^m\end{aligned}$$

当 $m<0$ 时

$$R[m]=\operatorname{Res}[\boldsymbol{S}(z^{-1})z^{-m-1};z=\rho]$$

上式成立是因为 $z=p^{-1}$ 是单位圆外的一个极点。

进一步地

$$\begin{aligned}\boldsymbol{S}(1/z)&=-2(\rho^{-1}-\rho)\frac{z^{-1}}{(z^{-1}-\rho)(z^{-1}-\rho^{-1})}\\&=-2(\rho^{-1}-\rho)\frac{z}{(z-\rho^{-1})(z-\rho)}\end{aligned}$$

由于式(A.3-6)的对称性，上面的结果显而易见。

$$\begin{aligned}\operatorname{Res}[\boldsymbol{S}(z^{-1})z^{-m-1};z=\rho]&=-2(\rho^{-1}-\rho)\left.\frac{z^{-m}(z-\rho)}{(z-\rho^{-1})(z-\rho)}\right|_{z=p}\\&=-2\frac{(\rho^{-1}-\rho)\rho^{-m}}{(\rho-\rho^{-1})}\\&=2\rho^{-m}\end{aligned}$$

整理已得到的表达式，得到最终结果

$$R[m]=2\rho^{|m|},\qquad -\infty<m<+\infty$$

A.4 数学归纳法[①]

概率论里面的许多证明都是通过数学归纳法得到的。数学归纳法可用来证明许多定理，尤其是用其他方法证明比较困难的时候。例如：如果说一个集合 S 包含的元素都是正整数，该如何来验证这个结论呢？我们可以列举出 $1\in S$，$2\in S$，$3\in S$ 等。但要用这种方法的话，就不可能在有限长的时间里来完成证明。在这种情况下，就可以采用数学归纳法的思想。

令集合 $\{\mathcal{C}_k\}$ 为无限长的序列，其中 $k\geqslant 1$。我们希望当所有的 $k\geqslant 1$ 时该命题都成立。在这里不需要一个一个地证明，我们遵循以下的步骤来证明：

(i)如果 $\mathcal{C}_1$ 为真。

(ii)对任意的 $k>1$，如果 $\mathcal{C}_k$ 为真可得到 $\mathcal{C}_{k+1}$ 也为真，那么对所有的 $k\geqslant 1$，$\mathcal{C}_k$ 为真都成立。

那么当我们采用数学归纳法时，只需要完成两个步骤即可(i 和 ii)。对于待证明的命题，把子命题 $\mathcal{C}_k$ 编好号以后，首先证明 $\mathcal{C}_1$ 为真。然后再试图证明第二个步骤也为真。首先对于任意一个 $k>1$，假设 $\mathcal{C}_k$ 为真，如果可以得到 $\mathcal{C}_{k+1}$ 也为真。命题得证。

例 A.4-1 (数学归纳法)试证明当 $0<a<b$ 时，对于所有的正整数 n，有 $a^k<b^k$ 成立。

证明 在此我们选用数学归纳法。命题的条件 $0<a<b$ 可以直接当成第一个子命题 $\mathcal{C}_1=\{a<b\}$ 成立，那么令 $\mathcal{C}_k$ 为第 k 个子命题，定义为 $\mathcal{C}_k\triangleq\{a^k<b^k\}$。在此假设 $\mathcal{C}_k$ 为真，对于这个 k 有 $a^k<b^k$ 成立。可得到 $a^{k+1}=a\times a^k<a\times b^k<b\times b^k=b^{k+1}$。因此 $\mathcal{C}_{k+1}$ 为真。根据数学归纳法的思想可知对于所有的 $\mathcal{C}_k$ 为真都成立，因此对于所有的正整数 k，有 $a^k<b^k$ 成立。

---

① 参见参考文献[A-4]。

参考文献

A-1 A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing.* Upper Saddle River, NJ: Prentice Hall, 1989, Chapter 3.

A-2 T. Kailath, *Linear Systems.* Upper Saddle River, NJ: Prentice Hall, 1980, pp. 161-166.

A-3 E. Hille, *Analytic Function Theory*, Vol. I. New York: Blaisdell, 1965, Chapters 7 and 9.

A-4 S. L. Salas and Einar Hille, *Calculus*, 3rd edition, New York: John Wiley & Sons, 1978.

附录 B　伽马函数和 δ 函数

B.1　伽马函数

伽马函数 $\Gamma(\alpha)$（α 为实数）是通过积分来定义的[1,2]

$$\Gamma(\alpha) \triangleq \int_0^{\infty} t^{\alpha-1}\mathrm{e}^{-t}\mathrm{d}t \tag{B.1-1}$$

其中 $\alpha > 0$。由式(B.1-1)可得 $\Gamma(1) = 1$。利用分部积分法，可得 $\Gamma(\alpha+1)$

$$\begin{aligned}\Gamma(\alpha+1) &= \int_0^{\infty} t^{\alpha}\mathrm{e}^{-t}\mathrm{d}t = -t^{\alpha}\mathrm{e}^{-t}\Big|_0^{\infty} + \alpha\int_0^{\infty} t^{\alpha-1}\mathrm{e}^{-t}\mathrm{d}t \\ &= \alpha\Gamma(\alpha)\end{aligned}$$

因此对于所有的正整数 k

$$\Gamma(k) = (k-1)!$$

对于两个整数之间的自变量，伽马函数采用平滑内插来求值。在 MATLAB 中相应的函数为 `gamma`。

注意有 $0! = 1$。在这里我们给出两个结论 $\Gamma(0.5) = \sqrt{\pi}$ 和 $\Gamma(1.5) = \sqrt{\pi}/2$。伽马函数有的时候被称为广义阶乘函数。

B.2　不完全伽马函数

（上）不完全伽马函数 $\Gamma(\alpha, x)$ 也是通过积分定义

$$\Gamma(\alpha,x) \triangleq \int_x^{\infty} t^{\alpha-1}\mathrm{e}^{-t}\mathrm{d}t$$

其中 $\alpha > 0$。（下）不完全伽马函数的定义为

$$\gamma(\alpha,x) = \int_0^{x} t^{\alpha}\mathrm{e}^{-t}\mathrm{d}t$$

本书中如果不做特别说明的话，不完全伽马函数都是指上不完全伽马函数。很显然，$\Gamma(\alpha, 0) = \Gamma(\alpha)$。当 $\alpha = k$ 为整数时，不完全伽马函数满足下面的序列[3,4]：

$$\Gamma(k,x) = (k-1)!\mathrm{e}^{-x}\sum_{l=0}^{k-1}\frac{x^l}{l!}$$

上式也可以写成：

$$\Gamma(k,x) = (k-1)\Gamma(\alpha-1) + x^{k-1}\mathrm{e}^{-x}$$

在 MATLAB 对应的函数为 `gammainc`。该函数在计算泊松随机变量的分布中起到非常关键的作用。

B.3 δ 函数

δ 函数 $\delta(x)$ 被定义为这样的一个"函数"：该函数值在 $x \neq 0$ 时为 0，在 $x=0$ 时为无穷大

$$\int_{-\infty}^{\infty} \delta(x)\mathrm{d}x = 1$$

关于 $\delta(x)$ 是否为一个"函数"，争论主要在于该函数的取值是无界的，或者在一段有限长时间内该信号的能量竟然为无穷大。另外一种关于 $\delta(x)$ 的定义是把它看成一种脉冲函数的极限情况。比如对于矩形窗

$$w\left(\frac{x}{b}\right) \triangleq \begin{cases} 1, & -b/2 \leqslant x \leqslant b/2 \\ 0, & \text{其他} \end{cases}$$

那么就可以把 $\delta(x)$ 定义为

$$\delta(x) \triangleq \lim_{a\to\infty} \{aw(ax)\}$$

另外一种关于 $\delta(x)$ 的定义也可以是如下的表达式：

$$\delta(x) \triangleq \lim_{a\to\infty} \{a\exp(-\pi a^2x^2)\}$$

矩形形状和高斯形状的脉冲函数如图 B.3-1 所示。其中 $aw(ax)$ 的导数是不连续的，而 $a\exp(-\pi a^2x^2)$ 具有连续的导数。其实这些脉冲函数的具体形状是不重要的，$\delta(x)$ 函数最关键的性质为：(1)单位面积；(2)当 $x \neq 0$ 时函数值要迅速降低为 0。

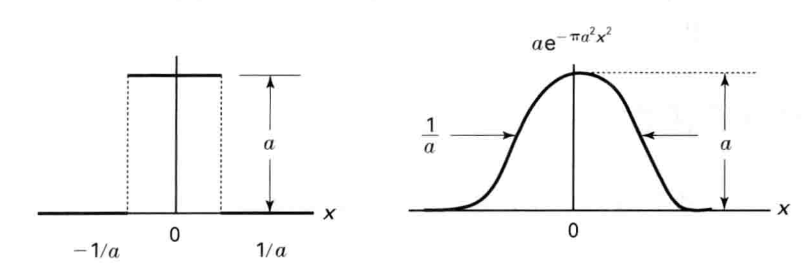


图 B.3-1 单位矩形和单位高斯形状脉冲

关于 δ 函数还可以给出这样的定义：如果任意一个函数 $f(\cdot)$ 关于 x 连续，且满足以下积分表达式，那么该函数就是 δ 函数①

$$\int_{-\infty}^{\infty} f(y)\delta(y-x)\,\mathrm{d}y = f(x) \qquad \text{(B.3-1)}$$

这里的定义和前面的那些定义其实是一致的，因为将 a 足够大的时候得到的 $\delta(x)$ 的定义代入到式(B.3-1)中也是成立的。由于只有在 $x \approx y$ 时候被积函数不为 0，因此可以用 $f(x)$ 来取代 $f(y)$，并把 $f(x)$ 移到积分符号外面去。由于所有的脉冲函数都是单位的，因此可以得到式(B.3-1)的结果。注意有 $\delta(x) = \delta(-x)$。

① 在此需要说明的是，由于 $\delta(x)$ 仅在一个孤立点上的取值不为 0，因此并没有其黎曼积分的定义。因此式(B.3-1)本质上只是一个符号表达式。换句话说，这只是对矩形脉冲和高斯形状脉冲的一种限制性操作。

对于单位阶跃函数 $u(x-x_i)$，该函数在 $x=x_i$ 是不连续的且有 $u(0)\triangleq 1$ [如图 B.3-2(a)所示]。这种不连续性可以看成图 B.3-2(b)中函数的极限情况，其导数如图 B.3-2(c)所示。

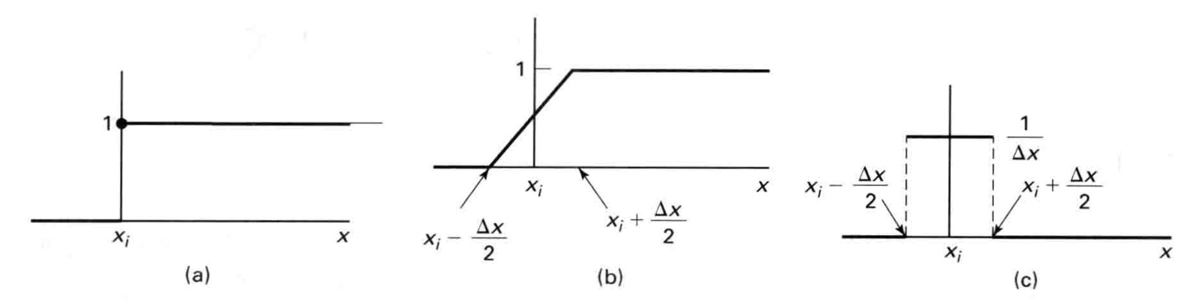


图 B.3-2　(a)单位阶跃函数 $u(x-x_i)$；(b)单位阶跃函数的一种近似；(c)图(b)中函数的导数

图 B.3-2(b)中函数的导数具体定义如下：

$$\left.\frac{\mathrm{d}F}{\mathrm{d}x}\right|_{x_i}\triangleq\frac{\mathrm{d}F(x_i)}{\mathrm{d}x_i}=\lim_{\Delta x\to 0}\frac{1}{\Delta x_i}\omega\left[\frac{x-x_i}{\Delta x}\right]=\delta(x-x_i)\tag{B.3-2}$$

因此，阶跃函数在非连续点处的导数就是一个 δ 函数，它的幅度①与阶跃的高度成正比。一般把 $\delta(x-x_i)$ 称为在 x_i 处的 δ 函数。

回到第 2 章中的式(2.5-7)，该公式可以写为

$$F(x)=\sum_i P_x(x_i)u(x-x_i)$$

再应用式(B.3-2)得到的结论，就可以写成一个离散随机变量的形式

$$f(x)=\frac{\mathrm{d}F(x)}{\mathrm{d}x}=\sum_i P_X(x_i)\delta(x-x_i)\tag{B.3-3}$$

其中定义 $P_X(x)\triangleq F(x_i)-F(x_i^-)$ 和单位阶跃函数保证了可以对所有的 i 进行求和操作，因此有 $x_i\leqslant x$。

参考文献

B-1　M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1972. (online at http://people. math. sfu. ca/ ~ cbm/aands/subj. htm

B-2　Gamma function Wikipedia page: http://en. wikipedia. org/wiki/Gamma function.

B-3　Incomplete Gamma function page at MathWorld: http://mathworld. wolfram. com/IncompleteGammaFunction. html.

B-4　Incomplete Gamma function Wikipedia page: http://en. wikipedia. org/wiki/Incomplete gamma function.

① 也可以称为 δ 函数的面积。

附录 C　函数变换与雅可比行列式

C.1　概要

函数变换在概率论中以及其他领域中的作用非常大。在本附录中我们将复习雅可比行列式的理论，首先复习两个函数之间的转换，然后再拓展到 n 个函数的情况。在此，首先复习高等数学中两个基本的结论。

定理 C.1-1　设 L 是从 E^n 到 E^n 的一个有界线性变换。如果 D 是 E^n 中的一个有界集合，且 D 的体积为 $V(D)$，那么 $L(D)$ 的体积为 $k\times V(D)$，其中 k 是独立于 D 的一个常数。

定理 C.1-2　如果 T 是开集 D 中从 E^n 到 E^n 的一个变换，且 $T\in C^1$，则对任意点 $p\in D$，$\mathrm{d}T$ 是 E^n 到 E^n 的一个线性变换。

第一个定理表明，变换 L 对集合 D 体积的影响只取决于一个常数，而与 D 的具体形状无关。第二个定理表明，在微分层面，即使非线性变换也成为线性的，这种变换具备了微分的功能。接下来将讲解这两个定理的具体应用。

C.2　$n=2$ 时的雅可比行列式

令 $v=g(x,\ y)$ 和 $w=h(x,\ y)$ 为一一映射[①]的微分方程对，它们各自的逆函数为 $x=\phi(v,\ w)$ 和 $y=\varphi(v,\ w)$。令向量 $\boldsymbol{z}=(v,\ w)$ 给出 v'-w' 平面上一个无限小的矩形 $\Re$，向量 $\boldsymbol{u}=(x,\ y)$ 也给出 x'-y' 平面上一个无限小的矩形 $\Im$。根据定理 C.1-1，这个微分变换是线性的，再根据定理 C.1-2，面积之比 $A(\Im)/A(\Re)$ 为一个常数。用 $|\tilde{J}|$ 来表示这个常数，接下来将计算 $|\tilde{J}|$ 的值。

在此可用图 C.2-1 来帮助计算常数 $\tilde{J}$ 的值。利用 $x=\phi(v,\ w)$ 和 $y=\varphi(v,\ w)$，顶点 $\boldsymbol{P}_1$，$\boldsymbol{P}_2$ 和 $\boldsymbol{P}_3$ 对应的映射点 $\tilde{\boldsymbol{P}}_1$，$\tilde{\boldsymbol{P}}_2$ 和 $\tilde{\boldsymbol{P}}_3$ 分别为

$$\tilde{\boldsymbol{P}}_1=(x,y),\ \tilde{\boldsymbol{P}}_2=\left(x+\frac{\partial\phi}{\partial v}\mathrm{d}v,y+\frac{\partial\varphi}{\partial v}\mathrm{d}v\right),\ \tilde{\boldsymbol{P}}_3=\left(x+\frac{\partial\phi}{\partial w}\mathrm{d}w,y+\frac{\partial\varphi}{\partial w}\mathrm{d}w\right)$$

直接将泰勒级数在 $(x,\ y)$ 处展开即可得到上面的结果。比如，对于 $\tilde{\boldsymbol{P}}_2$ 的坐标 $(x_2,\ y_2)$，计算过程如下：

$$x_2=\phi(v+\mathrm{d}v,w)\approx\phi(v,w)+\frac{\partial\phi}{\partial v}\mathrm{d}v\ \ 和\ \ y_2=\varphi(v+\mathrm{d}v,w)\approx\varphi(v,w)+\frac{\partial\varphi}{\partial v}\mathrm{d}v$$

关于 w 的导数都为 0，这是因为从 $\boldsymbol{P}_1$ 到 $\boldsymbol{P}_2$ 时 w 为一个常数。由向量分析的理论，向量 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 张成的平行四边形面积可由二者长度的叉乘给出

---

① 对于每个点 $(x,\ y)$，只唯一映射到点 $(w,\ v)$，反之亦然。

$$A(\Im)=|\boldsymbol{v}_1\times\boldsymbol{v}_2|=\left|\left(\frac{\partial\phi}{\partial v}\boldsymbol{i}\mathrm{d}v+\frac{\partial\varphi}{\partial v}\boldsymbol{j}\mathrm{d}v\right)\times\left(\frac{\partial\phi}{\partial w}\boldsymbol{i}\mathrm{d}w+\frac{\partial\varphi}{\partial w}\boldsymbol{j}\mathrm{d}w\right)\right|$$

其中 $\boldsymbol{v}_1=\tilde{\boldsymbol{P}}_2-\tilde{\boldsymbol{P}}_1$, $\boldsymbol{v}_2=\tilde{\boldsymbol{P}}_3-\tilde{\boldsymbol{P}}_1$。单位向量 $\boldsymbol{i}$ 和 $\boldsymbol{j}$ 满足 $\boldsymbol{i}\times\boldsymbol{j}=\boldsymbol{k}, \boldsymbol{j}\times\boldsymbol{i}=-\boldsymbol{k}, \boldsymbol{i}\times\boldsymbol{i}=\boldsymbol{j}\times\boldsymbol{j}=0$，其中 $\boldsymbol{k}\perp\boldsymbol{i},\boldsymbol{j}$，同时垂直于平面外的所有点。因此有

$$A(\Im)=\left|\frac{\partial\phi}{\partial v}\frac{\partial\varphi}{\partial w}-\frac{\partial\phi}{\partial w}\frac{\partial\varphi}{\partial v}\right|\mathrm{d}v\,\mathrm{d}w$$

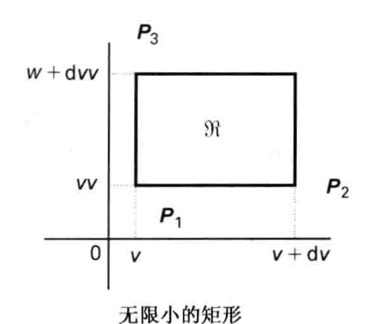
因为有 $A(\Re)=\mathrm{d}v\mathrm{d}w$，因此二者面积之比为

$$A(\Im)/A(\Re)=\left|\frac{\partial\phi}{\partial v}\frac{\partial\varphi}{\partial w}-\frac{\partial\phi}{\partial w}\frac{\partial\varphi}{\partial v}\right|\triangleq|\tilde{J}|$$

对于高维的情况，把 $\tilde{J}$ 写成行列式的形式更简洁一些。实际上，对于二维的情况也可以写成行列式的形式

$$\tilde{J}=\begin{vmatrix}\dfrac{\partial\phi}{\partial v} & \dfrac{\partial\phi}{\partial w}\\ \dfrac{\partial\varphi}{\partial v} & \dfrac{\partial\varphi}{\partial w}\end{vmatrix}=\frac{\partial\phi}{\partial v}\frac{\partial\varphi}{\partial w}-\frac{\partial\phi}{\partial w}\frac{\partial\varphi}{\partial v}$$

其中 $\tilde{J}$ 称为 $x=\phi(v,w)$ 和 $y=\varphi(v,w)$ 变换的雅可比行列式。



无限小的矩形

无限小的矩形映射为无限小的平行四边形

图 C. 2-1

除此之外，在计算概率测度（有时也称为概率质量或概率体积）的时候常常需要用到雅可比行列式。例如，假设概率密度函数为 $f_{XY}(x,y)$，变换关系为 $x=\phi(v,w)$ 和 $y=\varphi(v,w)$，定义事件 $B\triangleq\{\zeta:(X,Y)\in\wp\subset E^2\}$，那么有

$$P(B)=\iint_{\wp}f_{XY}(x,y)\mathrm{d}x\,\mathrm{d}y\neq\iint_{\wp}f_{XY}(\phi(v,w),\varphi(v,w))\mathrm{d}v\,\mathrm{d}w$$

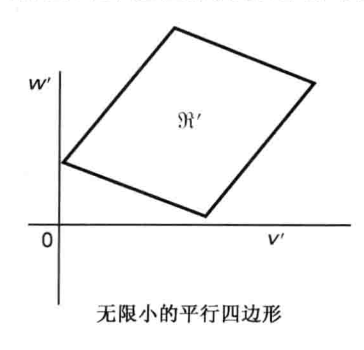
这是因为 $\mathrm{d}x\mathrm{d}x\neq\mathrm{d}v\mathrm{d}w$。这个时候就需要用雅可比行列式来让两个积分值之间划上等号，如下：

$$\iint_{\wp}f_{XY}(x,y)\mathrm{d}x\,\mathrm{d}y=\iint_{\wp}f_{XY}(\phi(v,w),\varphi(v,w))|\tilde{J}|\mathrm{d}v\,\mathrm{d}w$$

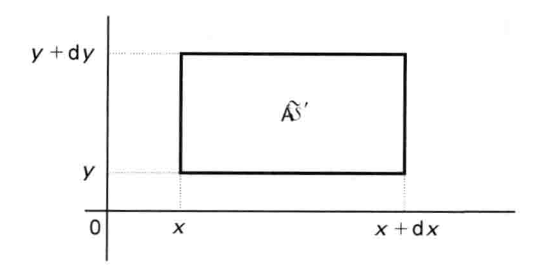
有的时候，对原函数（$v=g(x,y)$，$w=h(x,y)$）的操作比逆函数（$x=\phi(v,w)$，$y=\varphi(v,w)$）容易一些。为了得到想要的结果，我们需要计算面积之比，即 $v-w$ 系统中 $\Re'$ 比 $x-y$ 系统中的 $\Im'$（如图 C. 2-2 所示）。按照同样的步骤，可以得到 $A(\Im')/A(\Re')\triangleq1/|J|$，右上方的一撇用来区分两个系统中的区域。$J$ 的表达式如下：

$$J=\begin{vmatrix}\dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y}\\ \dfrac{\partial h}{\partial x} & \dfrac{\partial h}{\partial y}\end{vmatrix}$$

根据定理 C.1-1 可知 $A(\Im')/A(\Re')=A(\Im)/A(\Re)$，因此有 $|\tilde{J}|=1/|J|$，或者 $|\tilde{J}J|=1$ 成立。具体的计算过程作为课后练习留给读者。



无限小的平行四边形



无限小的平行四边形映射为无限小的矩形

图 C.2-2

C.3 n 为任意值时的雅可比行列式

如果我们愿意采用线性代数中的矩阵和向量符号，那么就更容易应对一般情况。首先，如果在高维情况下仍然采用单位向量 $\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}$ 非常不方便。此时，改为采用列向量表示的单位矩阵。因此在空间 $\boldsymbol{E}^2$ 中，有 $\boldsymbol{e}_1=[1,0]^{\mathrm{T}}$ 和 $\boldsymbol{e}_2=[0,1]^{\mathrm{T}}$。那么有

$$\boldsymbol{v}_1=\frac{\partial\phi}{\partial v}\mathrm{d}v\,\boldsymbol{e}_1+\frac{\partial\varphi}{\partial v}\mathrm{d}v\,\boldsymbol{e}_2=\left[\frac{\partial\phi}{\partial v}\mathrm{d}v,\frac{\partial\varphi}{\partial v}\mathrm{d}v\right]^{\mathrm{T}}$$

以及

$$\boldsymbol{v}_2=\frac{\partial\phi}{\partial w}\mathrm{d}w\,\boldsymbol{e}_1+\frac{\partial\varphi}{\partial w}\mathrm{d}w\,\boldsymbol{e}_2=\left[\frac{\partial\phi}{\partial w}\mathrm{d}w,\frac{\partial\varphi}{\partial w}\mathrm{d}w\right]^{\mathrm{T}}$$

接下来可得到 2×2 的矩阵 $\boldsymbol{V}_2=[\boldsymbol{v}_1\quad\boldsymbol{v}_2]$，其中 $\boldsymbol{V}_2$ 的下标 2 表示二维欧氏空间。

对于 $n=2$ 的这种特殊情况，$A(\Im)$ 的值即为 $|\det\boldsymbol{V}_2|$。当处理高维情况时，我们不再采用“平行多面体(六面体)面积”这个说法，而改为采用“平行多面体体积”，尽管此时还会有些较真的人坚持认为三维空间以上的情况应该采用“超体积”这个名词。同样地，在高维情况下，用不同的下标来区分函数和参量比用符号简单。在 n 维空间中，平行多面体体积仍然采用高乘以底面积的方法，其中底面积是指 $n-1$ 维空间中平行多面体的体积，高是指分量 $\boldsymbol{v}_n$ 的长度，而且 $\boldsymbol{v}_n$ 垂直于空间 E^{n-1} 中的所有向量。因此在空间 E^2 中，底面积是指选定的基向量的长度，高度是指第二个向量中与之正交分量的长度。在空间 E^3 中，底面积是指任意两个向量围成的平行四边形面积，高是指第三个向量与平面相垂直分量的长度，该平面是由前面两个向量构成的。

我们希望计算 n 维空间中无限小平行多面体的体积。在二维空间中，体积 $\boldsymbol{V}_2=|\det\boldsymbol{V}_2|$，那么在 n 维空间中直接得出 $\boldsymbol{V}_n=|\det\boldsymbol{V}_n|$，这样对吗？回答是“对”，而且可以由数学归纳法证明。在此，我们假设 $\boldsymbol{V}_n=|\det\boldsymbol{V}_n|$ 是正确的，那么就必须要证明 $\boldsymbol{V}_{n+1}=|\det\boldsymbol{V}_{n+1}|$ 也是正

确的。根据向量 $\boldsymbol{v}_1$，$\boldsymbol{v}_2$，…，$\boldsymbol{v}_n$，$\boldsymbol{v}_{n+1}$，矩阵 $\boldsymbol{V}_{n+1}$ 可以写为

$$\boldsymbol{V}_{n+1}=\left[\begin{array}{ccc|c}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n & v_{n+1,1}\\ \vdots & \vdots & \vdots & v_{n+1,2}\\ \vdots & \vdots & \vdots & \vdots\\ \hline 0 & 0 & 0 & v_{n+1,n+1}\end{array}\right]$$

$$=\left[\begin{array}{cc|c}\boldsymbol{V}_n & \begin{array}{c}v_{n+1,1}\\ \vdots\end{array} & \\ \hline 0\ \cdots & 0 & v_{n+1,n+1}\end{array}\right]$$

为了计算 $|\det \boldsymbol{V}_{n+1}|$，可将最下面一行展开，由于展开后其余部分都为 0，因此可得到 $|\det \boldsymbol{V}_{n+1}|=|v_{n+1,v+1}||\det \boldsymbol{V}_n|$。现在仔细分析一下向量 $\boldsymbol{v}_{n+1}$，根据向量 $\boldsymbol{e}_1$，$\boldsymbol{e}_2$，…，$\boldsymbol{e}_{n+1}$，它可以写为

$$\boldsymbol{v}_{n+1}=v_{n+1,n+1}\boldsymbol{e}_{n+1}+\sum_{i=1}^{n}v_{n+1,i}\,\boldsymbol{e}_i$$

其中 $\boldsymbol{e}_i$ 为一个列向量，其中第 i 行为 1，其余都为 0。$\boldsymbol{e}_{n+1}$ 为单位向量，与 $\boldsymbol{e}_1$，$\boldsymbol{e}_2$，…，$\boldsymbol{e}_n$ 都正交，因此正交于 $\boldsymbol{e}_1$，$\boldsymbol{e}_2$，…，$\boldsymbol{e}_n$ 张成的空间，而且它的高度为 $|v_{n+1,n+1}|$。由于 $|\det \boldsymbol{V}_n|$ 表示 n 维空间中平行多面体的体积，因此它是 $n+1$ 维空间中的底面积。所以表达式 $|\det \boldsymbol{V}_{n+1}|=|v_{n+1,n+1}||\det \boldsymbol{V}_n|$，其实也是高乘以底面积的形式，命题得证。

读者如果对阿达马(Hadamard)不等式和格拉姆-施密特(Gram-Schmidt)正交化过程熟悉的话，那么就可以不用数学归纳法而直接给出更快的证明，不过这样就不够直观易懂。

例 C.3-1　在第 5 章中我们考虑了如下的变换：

$$\begin{aligned}y_1&=g_1(x_1,x_2,\cdots,x_n)\\ y_2&=g_2(x_1,x_2,\cdots,x_n)\\ &\vdots\\ y_n&=g_n(x_1,x_2,\cdots,x_n)\end{aligned}$$

其中存在唯一的逆变换

$$\begin{aligned}x_1&=\phi_1(y_1,y_2,\cdots,y_n)\\ x_2&=\phi_2(y_1,y_2,\cdots,y_n)\\ &\vdots\\ x_n&=\phi_n(y_1,y_2,\cdots,y_n)\end{aligned}$$

此时，(y_1，y_2，…，y_n) 系统中矩形平行多面体映射成为 (x_1，x_2，…，x_n) 系统中的平行多面体，其中前者的体积为 $\prod_{i=1}^{n}|\mathrm{d}y_i|$，后者的体积为 $|\det \boldsymbol{V}_n|=|\det[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n]|$。通过对变换求微分，可以得到 $\boldsymbol{v}_i$，$i=1$，…，n 如下：

$$\boldsymbol{v}_i=\left[\frac{\partial\phi_1}{\partial y_i}\mathrm{d}y_i,\cdots,\frac{\partial\phi_n}{\partial y_i}\mathrm{d}y_i\right]^{\mathrm{T}},\qquad i=1,\cdots,n$$

附录 D　测度和概率

D.1　内容简介与基本概念

有些数学家把概率论看成测度论(measure theory)的一种特例。实际上，随机变量是一种可测函数；分布函数是一种测量；事件是可测集；样本描述空间和事件域统称可测空间；概率空间是一种测度空间。对于那些对测度理论基本概念不太熟悉的读者，我们在本附录中会给出一些结论。在此认为读者已经阅读过第 1 章的内容，而且对集合运算，域和 σ 域（σ-field)等概念比较熟悉。本附录中的内容改编自 Billingsley 的经典著作①。

设 Ω 为一个空间(全集)，A，B，C，$\cdots$ 是 Ω 的元素(子集)。用 ϕ 来表示空集。令 $\Im$ 为 Ω 上的一个集合域。如果 $\Im$ 是 Ω 上的一个 σ 域，那么 $(\Omega, \Im)$ 就是一个可测空间。令 μ 为 $\Im$ 上的一个集合函数②。如果 μ 满足下面的条件，那么 μ 就是一个测度：

(i) 令 $A \in \Im$，那么有 $\mu[A] \in [0, \infty)$。

(ii) $\mu[\phi] = 0$。

(iii) 如果 A_1，A_2，$\cdots$ 是 $\Im$ 中的不相交序列集合，且有 $\cup_{k=1}^{\infty} A_k \in \Im$，那么

$$\mu\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} \mu[A_k]$$

这个性质称为可列可加性。如果 $\mu[\Omega] < \infty$，测度 μ 为有限值，当 $\mu[\Omega] = \infty$ 时 μ 为无限值。当 $\mu[\Omega] = 1$ 时 μ 成为概率测度，参见第 1 章相关内容。如果 $\Im$ 是 Ω 中的一个 σ 域，那么 $(\Omega, \Im, \mu)$ 是一个测度空间。

可列可加性意味着有限可加性，即

$$\mu\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} \mu[A_k]$$

此时要求这些集合不相交。如果 $A \subset B$ 的时候 $\mu[A] \leqslant \mu[B]$ 成立，那么测度 μ 为单调的。这个证明比较容易，按照参考文献中测度理论的通用写法，$BA^c \triangleq B - A$，$B = (B - A) \cup AB = (B - A) \cup A$。由于 A 和 $B - A$ 是不相交的，因此可以得到 $\mu[B] = \mu[B - A] + \mu[A] \geqslant \mu[A]$。同样地，由于 $A \cup B = (A - B) \cup (B - A) \cup AB$，可以得到 $\mu[A \cup B] = \mu[A - B] + \mu[B - A] + \mu[AB]$。这个结论可以推广到 σ 域中的多个集合（σ 域中的多个集合称为 σ 集合)，可以得到

$$\mu\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} \mu[A_k] - \sum_{i<j} \mu[A_i A_j] + \cdots + (-1)^{n+1} \mu[A_1 A \cdots A_n]$$

---

① Patrick Billingsley, *Probability and Measure*. New York: John Wiley & Sons, 1978.

② 集合函数是一个实值函数，定义在空间 Ω 的子集域 $\Im$ 上。

当然了，只有这些集合测度值有限时上面的等式才成立。同样易知，$\mu[\cdot]$ 具有次可加性的性质

$$\mu\left[\bigcup_{k=1}^{n} A_k\right] \leqslant \sum_{k=1}^{n} \mu[A_k]$$

例 D.1-1 (Borel)　勒贝格(Lebesgue)测度。设 $\mathfrak{I}$ 为区间 $\Omega = (0, 1)$ 上的 σ 域。$\mathfrak{I}$ 中的元素被称为线性博雷尔集，该区间上的 σ 域被称为博雷尔域。对于实数直线上的 σ 域我们都采用这种记号。关于 $\mathfrak{I}$ 的测度 $\mu[\cdot]$ 定义为 $\mu = \lambda(a, b) \triangleq b - a$，其中 $b \geqslant a$。这种测度称为区间 $(a, b]$ 上的勒贝格测度。这个定义可直接用于实数直线 R^1。k 维欧氏空间中的勒贝格测度为

$$\mu = \lambda_k[x : a_i < x_i \leqslant b_i, i = 1, \cdots, k] \triangleq \prod_{i=1}^{k}(b_i - a_i)$$

因此，勒贝格测度长度可以是长度($k=1$)，面积($k=2$)，体积($k=3$)以及超体积($k>3$)。我们用 $\mathscr{B}^k$ 来表示这些广义矩形对应的 σ 域。

关于测度有许多重要的定理，我们列举部分在此。

定理 D.1-1　(平移不变性)令 $A \in \mathscr{B}^k$，$A + x \triangleq \{a + x: a \in A\}$。那么对于 x 的所有平移都有 $\lambda_k(A + x) = \lambda_k(A)$。

定理 D.1-2　(平移的勒贝格测度)令 $T: R^k \to R^k$ 表示欧氏空间中从 R^k 到 R^k 的平移，该平移为线性且非奇异。那么当 $A \in \mathscr{B}^k$ 时有 $TA \in \mathscr{B}^k$ 和 $\lambda_k(TA) = |\det T| \lambda_k(A)$ 成立。例如，如果 T 表示一个旋转或者反射，即正交变换或归一化，那么有 $|\det T| = 1$ 和 $\lambda_k(TA) = \lambda_k(A)$ 成立。

定理 D.1-3　(R^k 子空间的勒贝格测度)每个 $k-1$ 维超平面的 k 维勒贝格测度为 0。

定理 D.1-4　(测度的连续性)(i)令 μ 为域 $\mathfrak{I}$ 的测度。如果 A_n 和 A 位于 $\mathfrak{I}$ 之内且 $A_n \uparrow A$，那么 $\mu[A_n] \uparrow \mu[A]$。这就称为测度的自下连续性。其中 $A_n \uparrow A$ 表示 $A_{n-1} \subset A_n \subset A_{n+1} \subset \cdots$，且有

$$A = \bigcup_{n=1}^{\infty} A_n$$

同样地，$\mu[A_n] \uparrow \mu[A]$ 表示 $\mu[A_n] \leqslant \mu[A_{n+1}] \leqslant \mu[A]$，且 $\lim_{n\to\infty} \mu[A_n] = \mu[A]$。

(ii)令 μ 为域 $\mathfrak{I}$ 的测度。如果 A_n 和 A 位于 $\mathfrak{I}$ 之内且 $A_n \downarrow A$，那么 $\mu[A_n] \downarrow \mu[A]$。这就称为测度的自上连续性。其中 $A_n \downarrow A$ 表示 $A_{n-1} \supset A_n \supset A_{n+1} \supset \cdots$，且有

$$A = \bigcap_{n=1}^{\infty} A_n$$

同样地，$\mu[A_n] \downarrow \mu[A]$ 表示 $\mu[A_n] \geqslant \mu[A_{n+1}] \geqslant \mu[A]$，且 $\lim_{n\to\infty} \mu[A_n] = \mu[A]$。

测度的映射和函数

令 $(\Omega, \mathfrak{I})$ 和 $(\Omega', \mathfrak{I}')$ 分别为集合 $A \in \mathfrak{I}$ 和 $A' \in \mathfrak{I}'$ 中的两个可测空间。对于映射 $T: \Omega \to \Omega'$，其逆像为 $T^{-1}A' = \{\omega \in \Omega: T\omega \in A'\}$，其中 $A' \subset \Omega'$。如果对于任意 $A' \in \mathfrak{I}'$ 都有 $T^{-1}A' \in \mathfrak{I}$，那么这个映射就是可测的。例如，对于单位区间 $\Omega = (0, 1)$，$\mathfrak{I} = \mathscr{B}$，映射为 $Tx = x^2$。

此时 $\Omega = \Omega'$ 和 $\Im' = \mathscr{B}$。很显然，Ω' 中每个博雷尔区间的逆像都是 Ω 中的博雷尔区间。因此，T 是一个可测映射。

对于 Ω 的一个实函数 X，其像空间为 R^1。对于每个 $B \in \Im$，如果 X 的逆像 $X^{-1}B = \{\omega: X(\omega) \in B\} \in \Im$ 都成立，那么 X 是可测的。

D.2 测度理论中关于概率的应用

如果满足以下条件，σ 域 $\Im$ 上的集合函数 P 是一个概率测度：

(i) 对于每个 $A \in \Im$，$0 \leqslant P[A] \leqslant 1$。

(ii) $P[\phi] = 0$，$P[\Omega] = 1$。

(iii) 如果 $A_1, A_2, \cdots, A_k, \cdots$ 是 $\Im$ 集合中不相交的序列，那么

$$\bigcup_{k=1}^{\infty} A_k \in \Im$$

此时有：

$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]$$

(这就是概率测度的可列可加性质。)

分布测度

为了与本书上下文保持一致，在此用 ζ 取代 ω 来表示 Ω 中的元素。在本书中，一般把 ω 用于傅里叶变换的变量。令 $B \in \mathscr{B}$，博雷尔 σ 域是实数直线上的区间。定义 $(R^1, \mathscr{B})$ 上的(概率)测度 μ 为 $\mu[B] \triangleq P[\zeta: X(\zeta) \in B] = P_X[B]$。这个测度称为随机变量的分布或规律。$X$ 的分布函数定义如下：

$$F_X(x) \triangleq \mu(-\infty, x] = P[X \leqslant x]$$

其中 $P[X \leqslant x]$ 是 $P[\zeta: X(\zeta) \leqslant x]$ 的简写。根据测度理论中的自上连续性，方程 $F_X(x)$ 是右连续的。

因为事件的域是一个 σ 域，分布方程是由测度给出的，所以测度的所有性质在概率论中也是成立的。因此，概率论和测度理论是紧密相关的。另外，通过测度理论的观点来认识概率论，就要忽略那些丰富的微积分理论，这些理论可以用来求解工程上的，科学上的和统计学上的问题。

附录 E　对模拟波形的采样和离散时间信号

离散时间信号通常是对连续时间的模拟波形采样得到的。在这里我们简要复习一下这两种信号的关系。利用均匀采样的信号来重建连续时间信号，这个重建的过程要遵循著名的奈奎斯特(Whittaker-Nyquist-Shannon)采样定理。

定理 E.1-1　一个最高实频率不超过 $v_{\max}$ 的连续时间信号 $x(t)$，如果采样间隔 T 满足 $T<\dfrac{1}{2v_{\max}}$，那么就可以通过样本 $x(nT)$ 来精确重建信号 $x(t)$。

很多地方都给出了这个重要定理的证明，例如 1965 年 John Wiley & Sons 出版社出版的，John M. Wozencraft 和 Irwin M. Jacobs 所著的 *Principles of Communication Engineering*。令 $x(t)$，$y(t)$ 和 $h(t)$ 分别表示一个线性时不变系统的输入信号，输出信号和冲激响应。设带宽 B(单位：Hz)大于系统中任意信号的带宽以及系统带宽，令 $\Delta \triangleq 1/(2B)$。为书写简洁，给出定义如下：

$$\operatorname{sinc}(x) \triangleq \frac{\sin \pi x}{\pi x}$$

一个线性时不变系统，其输入与输出之间的关系如下：

$$y(t)=\int_{-\infty}^{\infty} h(s)x(t-s)\mathrm{d}s$$

根据采样定理可得

$$y(t)=\sum_{l} y(l\Delta)\operatorname{sinc}(2B[t-l\Delta])$$

$$x(t)=\sum_{l} x(l\Delta)\operatorname{sinc}(2B[t-l\Delta])$$

$$h(t)=\sum_{l} h(l\Delta)\operatorname{sinc}(2B[t-l\Delta])$$

将上面三个表达式代入输入输出积分表达式，并用不同的下标来进行区分，可以得到 $t=l\Delta$ 时 $y(t)$ 的值

$$y(l\Delta)=\sum_{n}\sum_{m} h(n\Delta)x(m\Delta)I(l,m,n)$$

其中有

$$I(l,m,n)\triangleq\int_{-\infty}^{\infty}\operatorname{sinc}(2B[s-n\Delta])\operatorname{sinc}(2B[s-(l-m)\Delta])\mathrm{d}s=0$$

当 l，m 和 n 都为实整数且 $l-m\neq n$ 时，上式对于任意给定的 Δ 值都成立。因此我们可以得到如下重要的结论：

$$y(l\Delta)=\sum_{n} h(n\Delta)x\Big([l-n]\Delta\Big)\Delta$$

一般都会把因子 Δ 合并到 $h(n\Delta)$ 里。在计算机中，函数的样本值就是一组序列，比如

$y(l\Delta) \triangleq y[l]$, $x(l\Delta) \triangleq x[l]$, $h(n\Delta) \triangleq h[n]$。至此，可以得到

$$y[n] = \sum_{n} h[n]x[l-n]$$

上式就是一个离散卷积表达式。需要记住的是：对于离散信号的处理都是通过其样本实现的，然后再通过滤波重建一个模拟信号波形。

还有一点需要注意的是，序列 $\{x[n]\}$ 并不包含采样周期的信息。比如，对于正弦函数 $x(t) = A\cos(\omega_r t + \theta)$。如果采样时间 $t = n\Delta$, $n = \cdots, -2, -1, 0, 1, 2, \cdots$，那么得到的样本为 $x(n\Delta) = A\cos(n\Delta\omega_r + \theta) = A\cos(n\omega + \theta) \triangleq x[n]$，其中 $\omega \triangleq \Delta\omega_r$。"频率" ω 是一个没有量纲的单位，这类似于"时间" n 也是没有量纲的。如果要转换为模拟频率 ω_r（弧度/秒）或 v_r (Hz)，需要用到公式 $\omega_r = \omega\Delta$ 或 $v_r = v\Delta$。例如，对序列 $\{x[n]\}$ 进行傅里叶变换可以得到正弦曲线的频谱，其归一化频率 ω 位于区间 $[-\pi, \pi]$ 之内。如果转换到模拟的弧度频率上，那么这个频谱曲线就位于区间 $[-2\pi B, 2\pi B]$ 之内。

附录 F　正态随机变量样本均值和方差的独立性[①]

在我们所接触到的所有概率论和统计分布中，正态(高斯)分布无疑是最重要的额度。这有许多原因，但中心极限定理(Central Limit Theorem，CLT)是最主要的。CLT 是指在合理和现实的条件下，独立同分布随机变量之和趋向于正态分布。根据这个性质，当样本数量很大时，就可以运用中心极限定理来解决很多统计方面的问题。读过第 6 章和第 7 章的读者应该已经注意到我们用中心极限定理来解题，否则那些题目很难被解出。

还有其他许多原因来解释为什么正态分布在概率论和统计中扮演着如此重要的角色。对于一维正态概率密度函数，它的两个参数是代数无关的，即它们二者可以在各自的区间上任意取值而不会互相影响。均值 μ 可以在区间 $(-\infty,\infty)$ 上任意取值，方差 σ^2 可以在区间 $(0,\infty)$ 任意取值。这就使得我们可以设计一个特定均值和方差的正态随机数发生器。对于多维正态分布也是类似的。根据给定的数学期望矩阵 $\boldsymbol{\mu}$ 和方差阵 $\boldsymbol{K}$，我们可以设计相应的正态随机数发生器。正态分布的概率密度函数也具备完备性，该性质在寻找一组最优估计器的时候非常重要，即最小方差无偏估计器。

对于一个给定的正态分布，如何估计 μ 和 σ^2 这两个参数是中心问题。对于 $X:N(\mu,\sigma^2)$，假设有 n 个独立同分布(i. i. d.)的观测。我们用 $(1/n)\sum_{i=1}^{n}X_i$（样本均值）来估计均值 μ，用 $(1/(n-1))\left(\sum_{i=1}^{n}\left(X_i-(1/n)\sum_{j=1}^{n}X_j\right)^2\right)$ 或 $(1/n)\left(\sum_{i=1}^{n}\left(X_i-(1/n)\sum_{j=1}^{n}X_j\right)^2\right)$（样本方差）来估计方差 σ^2。注意到样本均值和样本方差都采用同样的数据。很显然，样本均值和样本方差都是统计独立的[②]。为了证明这个结论，我们将运用概率论中的一个定理：对于一个具有两个随机变量 V 和 W 的联合矩量生成函数 $M_{VW}(t_1,t_2)$，因子分别为 $M_V(t_1)$ 和 $M_W(t_2)$，那么 V 和 W 是独立的。这个结论是从例 4.7-1 中特征函数推导出来的，也就是估计矩量生成函数在 $t=\mathrm{j}\omega$ 时的值。

在这里我们感兴趣的两个随机变量是估计值 $\hat{\mu}_X$ 和 $\hat{\sigma}_X^2$。为行文简洁，给出如下定义：

$$\text{(i) } Y_i \triangleq \frac{X_i-\mu_X}{\sigma_X},\ \text{(ii) } V \triangleq n\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right)^2 = n\hat{\mu}_Y^2$$

$$\text{(iii) } W \triangleq \sum_{i=1}^{n}(Y_i-\hat{\mu}_Y)^2 = (n-1)\hat{\sigma}_Y^2$$

其中 $V:\chi_1^2$，$W:\chi_{n-1}^2$。关于 $M_{VW}(t_1,t_2)$，定义如下：

$$M_{VW}(t_1,t_2) = E\left[\exp(t_1V+t_2W)\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(2\pi)^{-n/2}\left[\exp\left(-\frac{1}{2}Q\right)\right]\mathrm{d}y_1\mathrm{d}y_2\cdots\mathrm{d}y_n$$

其中

---

① 严格的证明请参阅参考文献[7-1]。

② 独立或独立性都是统计意义上的。除此之外是指代数或方程的无关性。

$$Q \triangleq \sum_{i=1}^{n} y_i^2 - 2\frac{t_1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 - 2t_2\sum_{i=1}^{n}\left(y_i - \left(\frac{1}{n}\sum_{j=1}^{n} y_j\right)\right)^2$$
$$\triangleq \sum_{j=1}^{n}\sum_{i=1}^{n} r_{ij} y_i y_j = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{y}$$

在这里 $\boldsymbol{R}$ 是方差阵，它的对角元素是 r_{ii}，非对角元素为 r_{ij}，$i \neq j$ 具体为

$$r_{ii} = 1 - 2t_2 - \frac{2(t_1 - t_2)}{n}, i = 1, \cdots, n, \quad \boldsymbol{R} \text{ 的对角元素} \tag{F-1a}$$

$$r_{ij} = -\frac{2(t_1 - t_2)}{n}, i, j = 1, \cdots, n; i \neq j, \quad \boldsymbol{R} \text{ 的非对角元素} \tag{F-1b}$$

对于多维正态概率密度函数，其表达式如下：

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{R}|^{1/2}}\left[\exp(-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{y})\right]$$

其中 $\int_{-\infty}^{\infty} f_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = 1$。在此归纳如下：

$$M_{VW}(t_1, t_2) = E(\exp(t_1 V + t_2 W))$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(2\pi)^{-n/2}\left[\exp\left(-\frac{1}{2}Q\right)\right] \times \mathrm{d}y_1\mathrm{d}y_2\cdots\mathrm{d}y_n$$
$$= |\boldsymbol{R}|^{-1/2}$$

根据矩阵理论可知，对于 $n \times n$ 维的矩阵 $\boldsymbol{R}$，如果它的对角线元素为 a，非对角线元素为 b，那么该矩阵的行列式值 $|\boldsymbol{R}|$ 为 $(a-b)^{n-1}(a+(n-1)b)$。代入定义 $a \triangleq r_{ii}$，$b \triangleq r_{ij}$［根据式(F-1)］，可以得到

$$M_{VW}(t_1, t_2) = (1 - 2t_1)^{-1/2}(1 - 2t_2)^{-[(n-1)/2]}, t_1 < 1/2, t_2 < 1/2$$
$$= M_V(t_1) \times M_W(t_2)$$

根据前面引用的定理可知 V 和 W 是独立的。进一步有 $F_{VW}(v, w) = F_V(v)F_W(w)$，因此 $\hat{\mu}_X^2$ 和 $\hat{\sigma}_X^2$ 是相互独立的。同时，如果 $\hat{\mu}_X^2$ 和 σ_X^2 是相互独立的，那么 $\hat{\mu}_X$ 和 σ_X^2 也是相互独立的。这个重要的结论使得我们可以分别给 $\hat{\mu}_X$ 和 σ_X^2 选择不同的置信区间，而不必担心二者互相矛盾。只有在正态分布的时候，$\hat{\mu}_X$ 和 σ_X^2 才是相互独立的。

附录 G　累积分布函数查询表：正态分布，t(学生)分布，Chi 平方分布和 F 分布

在此我们给出 一些常见概率分布函数的查询表，包括：(1)正态分布；(2)学生 t 分布(t 分布)；(3)Chi 平方分布；(4)F 分布。F 分布有的时候称为 Snedecor F 分布。

伽马函数 $\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}\mathrm{e}^{-x}\mathrm{d}x$，$\alpha>0$ 在下面概率分布函数的定义中经常出现。当 α 为整数，或者说 $\alpha=m\geqslant 1$ 时，$\Gamma(m)=[m-1]!=(m-1)\times(m-2)\times\cdots\times 2\times 1$，其中 $0!=1$。下面给出每个概率分布函数的定义，如下所述。

(1)标准正态分布(在概率和统计中经常用到)

$$F_{SN}(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x$$

标准正态分布函数是均值 μ 和方差 σ^2 的双参量函数。

(2) t 分布(区间估计，正态总体均值的假设检验，检验 μ 是否等于 μ_0)

$$F_T(x;m)=K\int_{-\infty}^{z}\frac{1}{\sqrt{(1+(x^2/m)}}\mathrm{d}x,\quad K\triangleq\frac{\Gamma([m+1]/2)}{\Gamma(m/2)\times\sqrt{\pi m}}$$

t 分布函数是参量 m 的函数，其中 m 被称为自由度(degree of freedom，DOF)。t 分布是 F 分布的一种特例。

(3)Chi 平方分布(正态总体方差的置信区间，检验 σ^2 是否等于 σ_0^2，Pearson 适合度)

$$F_{\chi^2}(x;m)=K'\int_0^x y^{m/2-1}\exp\left(-\frac{y}{2}\right)\mathrm{d}y$$

$$K'\triangleq\frac{1}{2^{m/2}\Gamma(m/2)}$$

Chi 平方分布函数的参数 m 也称为自由度。

(4)Snedecor F 分布(广义释然比，检验 σ_1^2 与 σ_2^2 是否相等)

$$F_F(x;m,n)=K"\int_0^x y^{m/2-1}\times\left(1+\frac{my}{n}\right)^{-(m+n)/2}\mathrm{d}y$$

$$K"\triangleq\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{m/2}$$

Snedecor F 分布的两个参量 m 和 n 称为 F 分布的自由度。其中如果提及自由度，默认首先是指参量 m。

表 1 标准正态分布的累积分布函数查询表

根据 $F_{SN}(x)$ 来查询。x 的第一个数字给出表格的行，x 的第二个数字给出了该行中的位置

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

表2 t分布的累积分布函数查询表

$F_T(x;\ n)$ 的数值由表格顶部确定。首先根据 n 确定表格的行数，然后再查询相应的 x

F / n	0.60	0.75	0.90	0.95	0.975	0.99	0.995	0.9995
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.291

表中数据引自 W. H. Beyer, Ed., in *CRC Handbook of Tables for Probability and Statistics*, 2d ed., The Chemical Rubber Co., Cleveland, 1968; p. 283. 已获得 CRC 出版社的许可引用。

表3 Chi 平方分布的累积分布函数查询表

$F_{\chi^2}(x;\ n)$的数值由表格顶部确定。首先根据 n 确定表格的行数，然后再查询相应的 x

n/F	0.005	0.010	0.025	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.975	0.990	0.995
1	0.04393	0.03157	0.03982	0.02393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

该表引自文献 Tables of Percentage Points of the Incomplete Beta Function and of the Chi-square Distribution，"*Biometrika* Vol. 32 (1941)，并做了部分删减。已获得了作者 Catherine M. Thompson 和编辑 Biometrika 的许可引用。

表4 Snedecor F分布的累积分布函数查询表

查询表中函数F_F $(x;m,n)$三个参量的位置：左边第2列代表n，最上面一行代表m，最左边代表x

G	n m	1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	120	∞
0.90		39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.7	61.2	61.7	62.3	62.8	63.1	63.3
0.95		161	200	216	225	230	234	237	239	241	242	244	246	248	250	252	253	254
0.975	1	648	800	864	900	922	937	948	957	963	969	977	985	993	1000	1010	1010	1020
0.99		4,050	5,000	5,400	5,620	5,760	5,860	5,930	5,980	6,020	6,060	6,110	6,160	6,210	6,260	6,310	6,340	6,370
0.995		16,200	20,000	21,600	22,500	23,100	23,400	23,700	23,900	24,100	24,200	24,400	24,600	24,800	25,000	25,200	25,400	25,500
0.90		8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.46	9.47	9.48	9.49
0.95		18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
0.975	2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5
0.99		98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5
0.995		199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199
0.90		5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.17	5.15	5.14	5.13
0.95		10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.62	8.57	8.55	8.53
0.975	3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.3	14.2	14.1	14.0	13.9	13.9
0.99		34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.5	26.3	26.2	26.1
0.995		55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	43.4	43.1	42.8	42.5	42.1	42.0	41.8
0.90		4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.93	3.92	3.90	3.87	3.84	3.82	3.79	3.78	3.76
0.95		7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.75	5.69	5.66	5.63
0.975	4	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.46	8.36	8.31	8.26
0.99		21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.8	13.7	13.6	13.5
0.995		31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.7	20.4	20.2	19.9	19.6	19.5	19.3
0.90		4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.17	3.14	3.12	3.11
0.95		6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.50	4.43	4.40	4.37
0.975	5	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.23	6.12	6.07	6.02
0.99		16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.38	9.20	9.11	9.02
0.995		22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.4	13.1	12.9	12.7	12.4	12.3	12.1
0.90		3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.80	2.76	2.74	2.72
0.95		5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.81	3.74	3.70	3.67
0.975	6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.07	4.96	4.90	4.85
0.99		13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.23	7.06	6.97	6.88
0.995		18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.2	10.0	9.81	9.59	9.36	9.12	9.00	8.88

(续表)

G	n	m 1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	120	∞
0.90		3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.56	2.51	2.49	2.47
0.95		5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.38	3.30	3.27	3.23
0.975	7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.36	4.25	4.20	4.14
0.99		12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	5.99	5.82	5.74	5.65
0.995		16.2	12.4	10.9	10.1	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97	7.75	7.53	7.31	7.19	7.08
0.90		3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.38	2.34	2.31	2.29
0.95		5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.08	3.01	2.97	2.93
0.975	8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.89	3.78	3.73	3.67
0.99		11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.20	5.03	4.95	4.86
0.995		14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.40	6.18	6.06	5.95
0.90		3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.25	2.21	2.18	2.16
0.95		5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.86	2.79	2.75	2.71
0.975	9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.56	3.45	3.39	3.33
0.99		10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.65	4.48	4.40	4.31
0.995		13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03	5.83	5.62	5.41	5.30	5.19
0.90		3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.15	2.11	2.08	2.06
0.95		4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.84	2.77	2.70	2.62	2.58	2.54
0.975	10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.31	3.20	3.14	3.08
0.99		10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.25	4.08	4.00	3.91
0.995		12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47	5.27	5.07	4.86	4.75	4.64
0.90		3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.01	1.96	1.93	1.90
0.95		4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.47	2.38	2.34	2.30
0.975	12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	2.96	2.85	2.79	2.72
0.99		9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.70	3.54	3.45	3.36
0.995		11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72	4.53	4.33	4.12	4.01	3.90
0.90		3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.87	1.82	1.79	1.76
0.95		4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.25	2.16	2.11	2.07
0.975	15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.64	2.52	2.46	2.40
0.99		8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.21	3.05	2.96	2.87
0.995		10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07	3.88	3.69	3.48	3.37	3.26

(续表)

G	n	m	1	2	3	4	5	6	7	8	9	10	12	15	20	30	60	120	∞
0.90			2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.74	1.68	1.64	1.61
0.95			4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.04	1.95	1.90	1.84
0.975	20		5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.35	2.22	2.16	2.09
0.99			8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.78	2.61	2.52	2.42
0.995			9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50	3.32	3.12	2.92	2.81	2.69
0.90			2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.61	1.54	1.50	1.46
0.95			4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.84	1.74	1.68	1.62
0.975	30		5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.07	1.94	1.87	1.79
0.99			7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.39	2.21	2.11	2.01
0.995			9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01	2.82	2.63	2.42	2.30	2.18
0.90			2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.48	1.40	1.35	1.29
0.95			4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.65	1.53	1.47	1.39
0.975	60		5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.82	1.67	1.58	1.48
0.99			7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.03	1.84	1.73	1.60
0.995			8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57	2.39	2.19	1.96	1.83	1.69
0.90			2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.54	1.48	1.41	1.32	1.26	1.19
0.95			3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.55	1.43	1.35	1.25
0.975	120		5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.69	1.53	1.43	1.31
0.99			6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.86	1.66	1.53	1.38
0.995			8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37	2.19	1.98	1.75	1.61	1.43
0.90			2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.34	1.24	1.17	1.00
0.95			3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.46	1.32	1.22	1.00
0.975	∞		5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.57	1.39	1.27	1.00
0.99			6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.70	1.47	1.32	1.00

*该表格来自文献“Table of percentage points of the inverted beta distribution,” *Biometrika*, Vol. 33 (1943).，并做了部分删减。已获得了作者Maxine Merrington, Catherine M. Thompson和编辑Biometrika的许可引用。



.信息中心
北京市东城区北三环东路36号
北京环球贸易中心D座1208室
邮政编码：100013
电话：(8610) 57355171/57355169/57355176
传真：(8610) 58257961

Beijing Pearson Education
Information Centre
Suit 1208, Tower D, Beijing Global Trade Centre,
36 North Third Ring Road East,
Dongcheng District, Beijing, China 100013
TEL: (8610)57355171/57355169/57355176
FAX: (8610)58257961

尊敬的老师：

您好！

为了确保您及时有效地申请教辅资源，请您务必完整填写如下教辅申请表，加盖学院公章后将扫描件用电子邮件的形式发送给我们，我们将会在2-3个工作日内为您开通属于您个人的唯一账号以供您下载与教材配套的教师资源。

请填写所需教辅的开课信息：

采用教材			□中文版 □英文版 □双语版
作　者		**出版社**	
版　次		**ISBN**	
课程时间	始于　年　月　日	**学生人数**	
	止于　年　月　日	**学生年级**	□专科　□本科1/2年级 □研究生　□本科3/4年级

请填写您的个人信息：

学　校			
院系/专业			
姓　名		**职　称**	□助教 □讲师 □副教授 □教授
通信地址/邮编			
手　机		**电　话**	
传　真			
official email(必填) (eg:XXX@ruc.edu.cn)		**email (eg:XXX@163.com)**	
是否愿意接受我们定期的新书讯息通知：	□是　□否		

Publishing House of Electronics Industry
电子工业出版社：www.phei.com.cn
www.hxedu.com.cn
北京市万寿路173信箱高等教育分社（100036）
联系电话：010-88254555
E-mail：Te_service@phei.com.cn

系 / 院主任：________________（签字）

（系 / 院办公室章）

_____年_____月_____日